

大家心理學叢書／吳靜吉博士策劃

中央研究院副院長
曾志朗◎專文推薦

Keith E. Stanovich◎著
楊中芳等◎譯

這才是心理學！

教你如何以科學方法，培養批判性思考的能力

How to Think Straight about Psychology

大眾心理學叢書

吳靜吉博士策劃

每冊都解決一個或幾個你面臨的問題
每冊都包含可以面對問題的根本知識

300

這才是心理學！

大眾心理學叢書 300

這才是心理學！——教你如何以科學方法，培養批判性思考的能力

作　　者——Keith E. Stanovich
譯　　者——楊中芳等
策　　劃——吳靜吉博士
主　　編——林淑慎
特約編輯——陳錦輝
發 行 人——王榮文
出版發行——遠流出版事業股份有限公司
臺北市 100 南昌路二段 81 號 6 樓
郵撥／0189456-1
電話／2392-6899　　傳真／2392-6658
香港發行——遠流（香港）出版公司
香港北角英皇道 310 號雲華大廈 4 樓 505 室
電話／2508-9048　　傳真／2503-3258
香港售價／港幣 106 元
法律顧問——王秀哲律師・董安丹律師
著作權顧問——蕭雄淋律師
2005 年 4 月 1 日　初版一刷
行政院新聞局局版臺業字第 1295 號
售價新台幣 320 元（缺頁或破損的書，請寄回更換）

ISBN 957-32-5490-5　（英文版 ISBN 0-205-36093-9）
YLib 遠流博識網
http://www.ylib.com　E-mail:ylib@ylib.com

這才是心理學！

楊中芳等 譯

How to Think Straight about Psychology

Keith E. Stanovich

《大眾心理學叢書》出版緣起

王榮文

一九八四年，在當時一般讀者眼中，心理學還不是一個日常生活的閱讀類型，它還只是學院門牆內一個神秘的學科，就在歐威爾立下預言的一九八四年，我們大膽推出《大眾心理學全集》的系列叢書，企圖雄大地編輯各種心理學普及讀物，迄今已出版達二百種。

《大眾心理學全集》的出版，立刻就在台灣、香港得到旋風式的歡迎，翌年，論者更以「大眾心理學現象」為名，對這個社會反應多所論列。這個閱讀現象，一方面使遠流出版公司後來與大眾心理學有著密不可分的聯結印象，一方面也解釋了台灣社會在群體生活日趨複雜的背景下，人們如何透過心理學知識掌握發展的自我改良動機。

但十年過去，時代變了，出版任務也變了。儘管心理學的閱讀需求持續不衰，我們仍要虛心探問：今日中文世界讀者所要的心理學書籍，有沒有另一層次的發展？

在我們的想法裡，「大眾心理學」一詞其實包含了兩個內容：一是「心理學」，指出叢書的範圍，但我們採取了更寬廣的解釋，不僅包括西方學術主流的各種心理科學，也包括規

範性的東方心性之學。二是「大眾」，我們用它來描述這個叢書的「閱讀介面」，大眾，是一種語調，也是一種承諾（一種想為「共通讀者」服務的承諾）。

經過十年和二百種書，我們發現這兩個概念經得起考驗，甚至看來加倍清晰。但叢書要打交道的讀者組成變了，叢書內容取擇的理念也變了。

從讀者面來說，如今我們面對的讀者更加廣大、也更加精細（sophisticated）；這個叢書同時要了解高度都市化的香港、日趨多元的台灣，以及面臨巨大社會衝擊的中國沿海城市，顯然編輯工作是需要梳理更多更細微的層次，以滿足不同的社會情境。

從內容面來說，過去《大眾心理學全集》強調建立「自助諮詢系統」，並揭櫫「每冊都解決一個或幾個你面臨的問題」。如今「實用」這個概念必須有新的態度，一切知識終極都是實用的，而一切實用的卻都是有限的。這個叢書將在未來，使「實用的」能夠與時俱進（update），卻要容納更多「知識的」，使讀者可以在自身得到解決問題的力量。新的承諾因而改寫為「每冊都包含你可以面對一切問題的根本知識」。

在自助諮詢系統的建立，在編輯組織與學界連繫，我們更將求深、求廣，不改初衷。

這些想法，不一定明顯地表現在「新叢書」的外在，但它是編輯人與出版人的內在更新，叢書的精神也因而有了階段性的反省與更新，從更長的時間裡，請看我們的努力。

這才是心理學！

目錄

〈專文推薦〉

遠處的太陽正慢慢升起，照亮了我案頭的書！

曾志朗

我做記憶的研究，主要在探討人類如何把學會的新事物儲存在腦中、存在腦的哪一個部位？將來要用這些知識時，以哪一種提取的方式最爲快速而有效？爲什麼會有遺忘的現象？老人的記憶在哪一方面比年輕人差，爲什麼呢？是老化的生理因素造成過多的遺忘，還是老人們一生有太多的記憶造成提取時的干擾，所以對從前的經歷如數家珍，而對眼前經歷的事轉眼就忘？爲了研究這些問題，且能準確的找到答案，我必須有設備良好的實驗室，裡面有高速的電腦來幫助呈現學習材料，記錄被測試對象的反應，並根據不同實驗的目的，變化刺激材料呈現時的各種狀況。有時候，受試者還要戴上電極帽，讓我們能測量到腦波的變化；如果必要，受試者就會被帶到腦造影的實驗室，透過功能性磁共振造影（fMRI）儀器，把他們在學習以及在回憶時的腦部活動顯影在電腦螢幕上。

很多朋友來實驗室參觀，他們總是說：「你不是個心理學家嗎？幹嘛要這麼多設備去做實驗!?」有的人更語帶揶揄說：「佛洛伊德只有一張沙發就夠了！」碰到這種情況，我總要費很多口舌去說明「實驗」心理學的研究方法，並引經據典說明實驗心理學家這一百多年來的成就。但如果一開始

我的介紹是：「我們是認知神經科學家，在做記憶的研究。」則大多數的參訪者都會認為我們是很「科學」的一群，在做「很了不起」的實驗。詭異的是，我還是我，實驗室是同一個，設備儀器都沒有變，所做的實驗也相同，不一樣的只是「稱謂」而已。這是因為在一般人的心目中，心理學和科學是不搭調的，所以許多自稱為「心理學家」的議論是可以不必受到科學方法檢定的。但我要大聲疾呼的是，心理學眞的不是這樣的！在各種場合，我都一再強調心理學是一門很嚴謹的「人」的科學，而我常聽到的反應是：「你做的這些實驗當然很科學，但你不是心理學家，你是認知神經科學家啊！」

其實稱謂並不打緊，眞正的心理學在做什麼、怎麼做才最重要。多年來，我一直想寫一本科普書，把這一百多年來，心理學如何從沙發椅上的冥思走進實驗室的過程，做一個深入淺出的說明；並且把心理學自從進入實驗室後，如何對人的各方面行為有了新的了解，進而建立科學理論的成就，也做一些交代。但是很多事情不一定要自己才能做，很多時候就會有一些擁有共同理念的人搶先做了，而且做得比自己去做要好太多了。這本書就是一個最好的見證。作者 Keith Stanovich 是我多年的老朋友，我們同是美國心理協會會士，他在研究閱讀與推理歷程的成就非凡。二十年前，他和我都在幾個重要的科學期刊上當編輯委員，我們在各自的學校裡也都教一門「研究方法與設計」課，碰在一起時，總會交換課程大綱與教學心得，都感到大學生對心理學誤解重重的憂慮。這些誤解反映出整個社會大眾與媒體對科學心理學的無知，因無知加上誤解導致許多怪現象（如用指紋測

智力，常用左手可以開發右腦潛力等），也容易因無知引起的不安而被有心人操弄，造成許多人間悲劇，耗費社會成本。Keith說，他決心要寫一本書來「以正視聽」！他是劍及履及的人，所以，當我回台灣教書時，包裹裡就是他的第二版 *How to Think Straight about Psychology*。

這本書一出版，眞的是「轟動武林」，因爲幾乎所有大學裡教「普通心理學」的老師，都指定它爲學生必讀的補充材料，大家期待已久的一本「訓練心理學學生有批判性思考」的武功秘笈終於問世了！也因爲廣受歡迎，很快的在三年內更新增訂。

我帶著這本書，逢人必介紹，也用它來做爲普通心理學的教科書之一。十幾年來，台灣很多心理學界的朋友被我的熱情感染，也用它來做學生的必要讀物，但英文書總是普及得不夠快，當時從香港回台灣客座教書的楊中芳老師說她願意把它翻譯成中文，讓更多的人可以閱讀。但是楊老師回香港後，又到廣州的中山大學創設心理系，從各項實驗室的建立、人才的引進、教室的規劃，一切從無到有去建設，又要張羅行政的各項措施，還要籌措經費好讓年輕的老師得以發揮所長，自己的研究更不能丟掉，我看她非常忙碌，實在不忍催她。一直等到英文第七版問世，她由廣州傳來訊息，說譯稿已完成，要我爲文推薦。我突然驚醒，再忙碌的人，心中有理想，就會完成它。爲了這本書，她消瘦了，鬢髮亦轉蒼，但交稿的那天，我實在感動，她容光煥發，精神十足說：「喏，這才是心理學！」

我竟夜閱讀楊老師的譯稿，爲這本書寫推薦時，心裡不再惦著要去介紹書的內容，因爲只要讀

者捧起書，我相信他（她）一定會爲它精采的內容所吸引，讀完之後也一定會對「人」的複雜有所感觸，以後更會對心理學的各項報導，產生自發性的批判性思考。不論讀者是心理學或其他學科的專業，都將因爲這本書而開始對認眞研究眞正的心理學的人有所尊敬。人的實驗，遠比其他學科的實驗要難，因爲人的個別差異很大，周遭環境的少許風吹草動、鄰近社區的文化變異，或臨場身心狀況的變化，都會導致實驗結果的誤差，所以正確知識的建立很難。但多年來，研究者兢兢業業，在一個又一個嚴格控制的實驗之下，確實也取得了相當的成績，終於贏得其他學科的肯定與尊重。美國科學院二十幾年前才開始有心理學門的院士被選入，而中央研究院在成立將近六十年之後，我才被選進成爲第一位心理學的院士，可見一斑。

在整個人類知識的進展上，心理學必然會佔據越來越重要的位置，但什麼是心理學，一定要先弄清楚！我要向我的老朋友 Keith Stanovich 及他的書的譯者楊中芳老師致敬。啊！遠處的太陽正慢慢升起，照亮了我案頭的書！

【推薦者簡介】曾志朗，美國賓州州立大學心理學博士，一九九四年當選中央研究院院士。曾任教於俄亥俄州立大學、耶魯大學、加州大學柏克萊分校，並先後擔任中正大學社會科學院院長，陽明大學副校長、校長，教育部長等職，現任中央研究院副院長。著有《用心動腦話科學》、《人人都是科學人》（遠流）等書。

譯序

立意要翻譯這本書已經是差不多二十年前的事了，當時這本書才出了第一版，暑假我正在大陸教一些研究生「研究方法」的課，發現了這本好書，立刻鼓勵其中一個研究生把它翻譯出來，可惜後來因爲種種原因，不了了之了。像這世界上的許多事情一樣，熱情過後，這個意念很快就被許多其他更急迫、更切身的事情所淹沒。

直到二〇〇二年，我在中山大學再次教「研究方法」這門課，用本書的第六版作爲教科書，得到學生熱烈的回響，把它翻譯出來的念頭才再度浮現。底稿是由那年上這門課的四名博士生及八名碩士生起的。在范照及鄒智敏的協調下，經過若干遍的互校、再讀、重修（正值原書改版），才到了我的桌上。我在幾經艱難的行事曆安排下，在廣州、珠海、舊金山、香港、北京、台北、青島；在飯堂裡、書桌前、沙發上、病榻邊，斷斷續續地逐字完成整本書的審校工作。眞是非常辛苦，我老了！

雖然說想翻譯這本書已經有二十年，但是現在回想起來，當時如果眞的翻譯成書，影響力必定

不如現在，因為當時人們對心理學的認識或理解，以及大衆媒體對心理學的重視或炒作，都遠遠不及現在。書中涉及的那些利用心理學專家的虛名在電視、電台大放厥詞，或是利用僞科學的論述進行招搖撞騙的事例，當時在海峽兩岸都尙不普遍。在台灣，雖然大部分的大學都已有心理學系，心理學在各行業的應用卻才剛開始，大衆媒體也尙沒有發現它的「價值」。而在大陸，則僅只有寥寥幾個大學有心理學系，各大企業及一般老百姓還根本不知道有心理學這一玩意兒。在那樣的時空環境下，自然就不會對書中所述及的事情產生共鳴。

現在，則不同了，尤其是在台灣。

由於人們的物質生活富裕了，對生活品質的要求提高了，也感受到與外在物質的豐碩相比之下內心的空虛，因此精神及心理的需求被抬上了桌面。大家有錢也有閒去探討人生的意義，去理解自己及為自己的行為尋求解釋，以及想從心理的層面去追求更美好的生活——包括個人的滿足及人際的和諧。

台灣的大衆媒體二十年來由解禁到開放走得很快，即反映了人們的這種需求，也成為人們滿足這些心理需求最方便及廉價的工具。人們的心理需求在報紙的文章、電視的節目及書店的書架上看得最清楚。舉電視為例，任何一家付費有線電視系統，都提供超過一百台，除了教人賺錢的股市分析台及教人花錢的購物台之外，還包括了許多宏揚各種宗教、傳播各類知識，以及提供各型娛樂的電視台。雖然，大部分節目還是以如何改善自己的形象及生活品質為主要賣點，不過已有不少涉及

如何認識、理解及預測自己及他人的行為等這些與心理學息息相關的議題。很可惜的是，在這些節目中，有關人們的性格及行為之討論類往往是與算命理、看相、卜卦、看時辰、論星座、解生辰八字、排紫微斗數等等掛鉤，與正規的心理學則可謂風馬牛不相及。

本來，這也無所謂，我們做我們的心理學研究，他們搞他們的節目賺錢。然而，當大衆媒體在節目中散播這類與人們的心理及行為相關的訊息時，受衆們會把這些內容看成是正規心理學在研究及累積的知識，從而對之產生許多誤解，那心理學就遭了「池魚之殃」，而不得不「有所謂」，而要「有所為」了。例如，在一些電視節目中，宣揚減肥、整容就一定可以令自己內外「煥然一新」，改名換姓就可以改變自己的命運，生辰八字或星座決定了自己及他人的性格等等論調，早已經被認為是無須實徵驗證的「基本心理學常識」了。

這些「基本心理學常識」毫無科學依據，然而卻沒有人有興趣或任何一點好奇心去問：這些常識都是正確的、可以信賴的嗎？本書認為人們之所以會不對這些「常識」置疑，並用之如飴，是因為由大衆媒體所培養出的一套所謂的「偽科學信念系統」已經根植於人們心中，造成他們對心理學是什麼有很大的誤解。這一套信念系統的主要觀點是：心理學不是一門科學，也沒有什麼一定的標準來幫助我們評估心理學知識的好壞，因此，「任憑你做什麼都行」，都可以是心理學，甚至只是隨便說說、自成一理，只要能有人受用，也可以是正確的、好用的心理學知識。有了這樣的一套信念系統，「心理學」自然是比比皆是，唾手可得。

在這個「僞科學知識」氾濫的時代，心理學家自己有時也扮演了助紂爲虐的角色。透過大衆媒體的大力推廣，現在心理學已經成爲一種時尙。任何事件請一個專家從心理學的角度來點評幾句，已經變成能夠滿足人們好奇心及偷窺欲的賣點，從而形成有如八卦雜誌一樣的「固定銷路」。「媒體心理專家」也應運而生，他們之中並不都是受過心理學訓練的專家，很多是報紙專欄作者、小說家、社會／政治評論家、記者或其他社會知名人士。總之，知名度高及口才伶俐是主要的入選標準。唯一例外的是各個相關領域的大學教授，他們雖然知名度不是最高，口才不算最伶俐，但也是在經常被邀請之列，其中包括不少心理學系的教授們。

然而，當這些心理學專家被邀請上電視時，他們並沒有利用這些機會去把他們經年研究的成果公諸於世，藉以傳播正確的心理學知識給大衆；相反地，他們的大學教職或博士學位，只是增加了他們的「權威性」及「可信度」，讓他們對任何事或人雖只要侃侃地談出一番與自己的所知所學毫無關係的「道理」，就有人聽；對任何個人的任何心理問題，雖只要信口地講出一套教其「自助」的方法，就有人從。

透過行銷專家的包裝、媒體心理學家的助長，製造出不少暢銷的「心理學常識」，建構了一個心理學的文化產業，然而這一產業搶走的卻是不少算命、看手相、看星座的節目或人的生意！這反映出許多這些「心理學文化產業」的消費者（讀者、觀衆、聽衆等），其實認爲心理學與算命、看手相、看星座沒有太大差別，並不把它看成是一門嚴肅的科學知識。他們之所以會比較聽信心理學專

家的話，是因爲他們覺得心理學這一名詞沒那麼俗，上得了學術的殿堂，或者是誤以爲那些所謂的「心理學專家」所給的「心理常識」是有科學研究爲依據的，所以是比較可靠的。

本書清楚地告訴我們：目前在坊間販賣的心理學知識大多是「僞科學知識」，不是眞正的心理學，是出於對心理學的誤解。身爲心理學工作者，不能再坐視自己的研究領域遭人踐踏，必須讓人們看清楚心理學到底是什麼。

心理學是一門科學。心理學的知識是靠科學研究累積而來的。現在在大衆媒體上，由大部分「心理專家」所散播的「心理學常識」，都是專家自己「個人意見」的表達、「個別經驗」的總結或「臨床個案」的舉例。按本書所定下的「科學研究標準」，都是不合格的心理學知識，因此其訊息的品質並不高，可信度也並不會比算命的、看手相的、看星座的好到哪裡去。如果發現這些專家講的「道理」對你很有用，那很有可能只是「安慰劑效應」而已（本書會詳細介紹這一效應）。

爲此，本書可以說是一本引導消費者（想要應用心理學知識的讀者）如何評價心理學文化產品的購買指南，讓想要應用這些「心理學常識」的消費者知道，如何判斷哪些「道理」是有科學依據的，因此是可信的；而哪些是不值一文的「廢話」，只有娛樂價值。

從這個角度看，這本書是適合一般有興趣的普通大衆看的，帶有糾正人們心目中對「心理學」之誤解的意味，故將書名定爲《這才是心理學！》。

從另外一個角度看，這本書作爲心理學系二年級學生的第一本「研究方法」課的教科書，也是

再適當不過的了。因爲它貼近生活，學生可以在日常生活中找到實例，來討論如何可以做一些眞正「科學」的心理學研究，從而更理性地探索人類心理的奧祕。也因爲作者文筆簡潔但鋒利，詼諧但冷靜，邏輯性很強，說服力極高，可以給學生帶來很大的腦力激盪及急轉彎，增加他們的創新思考能力。

總之，《這才是心理學！》對揭去心理學之朦朧（或錯誤）面紗做了很大的貢獻，讓讀者看清心理學的眞面目。我強力推薦這本書給所有想知道心理學是怎麼一回事、心理學知識如何累積的讀者。它指出心理學雖是一門新興的科學，卻擁有一個非常古老的傳統，因此人們對它的誤解很深，以爲對心理學研究所累積的知識，可以不用「科學」的標準來衡量及評價。本書則告訴我們，判定心理學知識的好壞，是和其他的科學一樣，一定要用「科學」的標準。本書用了許多精彩的例子說明這些標準是什麼，以及要如何用之分辨眞正的科學心理學知識與迷惑大衆的僞科學心理常識。

本書除了初稿的譯者郭妙燕（序言）、程樂華（第一章）、蕭曉雲（第二章）、潘偉（第三章）、蕭崇好（第四章）、徐四華（第五章）、馬淑蕾（第六章）、鄒智敏（第七章）、范照（第八章）、郭妙燕（第九章）、胡艷華（第十章）、韋慶旺（第十一章）和林升棟（第十二章）之外，中山大學心理學系許多其他的研究生及本科生同學都參與了互校、閱讀、找錯等的工作，他們的努力讓這本書能夠譯得更準確、更順暢。因此，值得在這裡把他們的名字記下，一併表示感謝。

二〇〇一年級本科生：丁宇、馬春蓉、史松衢、孫靜、何妙玲、李潔、李天放、蕭淩、蕭廬

奇、余夢、張浩宇、鄒恆、鄭茵、胡姍姍、趙冀、黃德斌、黃樂全、謝志成、曾廣斌、曾勇軍、賴雲鵬。二〇〇三年級研究生：趙瑾東、何淑瑜、鄧詠喻、劉雪琴、陳淑明、董蓓、袁佾芸、章睿健、王亞、崔吉芳、崔麗弦、劉淑華、阮福金。二〇〇四級博士生：楊斌讓。

本書可以說是中山大學復系（二〇〇一年）以來，幾屆師生努力的一個成果。它對我本人來說意義也很重大，我之決定離開香港大學，轉到中山大學胼手胝足地去打拚，就是希望能讓更多的人看到心理學研究的眞正價値及意義。所以，在徵得所有譯者的同意下，本書台灣版的版稅將全數捐給「中山大學心理學系發展基金」，做爲該系未來學科建設之用。

最後，我要感謝遠流出版公司的編輯同仁，對本書台灣版的細心校閱及編輯，以及我的老友曾志朗教授的鼓勵（寫推薦）、審閱（特別是專有名詞的翻譯）及建議書名。沒有他們的幫助，本書的台灣版是無法順利付梓的。

楊中芳

二〇〇五年二月於台北

前言

在人類知識的滄海中，有這樣一套知識體系，它尚不為大多數人所知，卻關係到人類各式各樣的行為和意識，我們可以用它來解釋、預測和控制人的行為。懂得這套知識體系的人可以用它來分析其他人，並能夠更完整、更精確地理解決定他們行為和想法的眞正原因。可能令你驚訝的是，這套不為人知的知識體系，竟會是心理學！

當我說心理學仍不為人知時，你一定心想我是在開玩笑吧。因為現在書店裡充斥著大量標示心理學為主題的書籍，電視和廣播每天頻頻播放關於心理學的脫口秀，報紙和雜誌也開闢有心理學專欄，怎麼能說沒有人知道心理學呢？然而，就某種至關重要的意義而言，心理學確實仍是一個不為人知的知識領域。

儘管心理學似乎得到很多媒體的關注，但是這個知識體系的絕大部分內容仍不為大衆所知。那些透過大衆傳媒傳播的「心理學」知識，很大程度上只是一種幻象。很多人不知道他們在書店看到的大部分所謂心理學書籍，都是一些在心理學界根本站不住腳的人寫的；很多人也不知道多數在電

視上被貼上**心理學家**標籤的那些人，根本得不到美國心理學會（American Psychological Association, APA）和美國心理協會（American Psychological Society, APS）的認可；他們更不知道，大多數高曝光率的心理學「專家」，對心理學領域知識的累積未曾做過任何貢獻。

傳媒對「心理學」話題的短期炒作，不僅向大眾傳遞許多不準確的訊息，還模糊了人們對心理學領域中眞正的、正在發展的知識體系的認識。大眾不能肯定到底什麼是心理學，什麼不是，也不知道如何自主地評估媒體上流傳的關於人類行爲的言論。更麻煩的是：很多人利用大眾沒有評估能力和技術的弱點，或利用他們認爲心理學言論無從評估的錯誤觀念，正在謀取私利。這種認爲心理學言論無法評估的謬論——有時也被理解爲「心理學沒有行規」，因此可以百無禁忌地「任憑做什麼都行」的想法——在當下處處可見，對社會危害極大，本書將詳細討論。很多僞科學正是利用大眾的這些無知，大行其道，從中獲取暴利，搖身演變成各類價值上百萬美元的產業。人們並不知道許多僞科學（例如占星術、通靈手術、速讀、生物節律、潛意識自療錄音帶和通靈偵探等）宣稱的眞理，其實已證實是錯誤的。本書提及的這些僞科學行業，助長了媒體對科學做出聳人聽聞報導的趨勢。這種趨勢對心理學的危害遠甚於其他學科，理解其中的原因，是學習如何正視心理學的一個重要環節。

本書不是寫給即將成爲心理學的研究者看的，而是爲一個更大的讀者群——應用心理學資訊的人——寫的，本書適用於初學心理學的學生，也適用於透過大眾媒體對心理學知識有些了解、又想

知道如何評價這些資訊之合理性的廣大讀者。

本書不是一本介紹心理學的正規入門書，它沒有總結心理學領域已取得的一系列研究成果。實際上，要糾正大衆那些已經被傳媒誤導的對心理學的看法，單靠到大學選修一門心理學導論的課程可能是不夠的，也不是解決問題的最好辦法。因爲很多對心理學大有興趣的外行人，他們沒時間、沒錢或沒機會到大學參加這種正規的學習。更重要的原因是，作爲一名大學心理學課程的教師，我不得不承認我和我的同事在教心理學時，也常常沒能好好引導初學者對心理學這門科學有正確的理解。因爲初級水準的課程設計中，通常沒有包括對批判性分析思維技巧的訓練，而這正是本書的焦點所在。作爲教師，我們常常只會忙於將「研究發現」塞入教學「內容」。每次我們和學生討論到諸如媒體眼中的心理學等稍微偏離教學大綱的話題時，都會感到內疚，並開始擔心自己會不會因爲離題，而不能在學期結束前教完自己眞正想講授的內容。

讓我們看看目前現有的一般心理學導論教科書，它們很多都有六到八百頁那麼厚，其中參考了幾百個已出版著作中的研究。當然，材料如此豐富並沒什麼錯，至少它反映了心理學知識的不斷增長。但這種情況也會帶來一些負面效果，教師常常只忙於塞給學生一大堆理論、事實和實驗，而沒時間回答學生在學習過程中的提問，也來不及糾正他們的錯誤觀念。這主要因爲教師（包括入門類教科書的作者）想當然地認爲只要學生接觸夠多的心理學研究，自然就可以從中推斷出他們所提問題的答案。他們希望學生可以從對心理學各領域研究之討論中，自行挖掘出隱藏其中對各類問題的

答案；但是這個希望往往落空，因爲到這門課的最後復習階段——經常是學期結束前的學習輔導時段——教師才震驚和沮喪地發現，學生提出的一些問題及說法，應該在課程開始第一天就提出來討論的，而不是最後一天。比如有學生提出：「既然心理學實驗不同於實際生活，那麼這些實驗能告訴我們什麼呢？」「心理學可以像化學那樣成爲一門眞正的科學嗎？」「可是我在電視上聽一位臨床心理學家講的，正好與我們教科書上說的相反！」「我認爲這個理論毫無價值——因爲我弟弟的行爲和它說的截然相反。」「心理學不過是些一般常識而已。」「每個人都知道什麼是焦慮——爲什麼還要費工夫去定義它呢？」「心理學只不過是一些不同看法的匯總而已，不是嗎？」對很多學生而言，這些問題是沒辦法僅靠思考書中的內容找到答案的。在本書中，我將明確回答上述這些問題，並澄清這些問題背後所反映的種種困惑。

許多研究都已證實，心理學入門課程通常不能有效糾正初學者對心理學的衆多誤解（Best, 1982; Higbee & Clay, 1998; McCutcheon, Furnham, & Davis, 1993; Vaughan, 1977）。一位研究者說：「我不得不說，一般那些『導論』性質的課程，事實上對糾正學生的錯誤觀念作用不大。」（Vaughan, 1977, p. 140）他還進一步得出結論：「目前我們在學生中還沒有培養出一種批判性思考的氛圍，這種氛圍可以讓學生對書中沒有講清楚的地方置疑。」（p. 140）後面這一結論與本書撰寫的初衷吻合，心理學也許比其他科學更需要批判性思考來引導學生將科學以外的雜草剷除。這種批判性思考是學生欲獨立評價心理學資訊所必須掌握的！

即使若千年後學生不再記得心理學入門課程中的內容，但仍然可以運用本書涵蓋的基本原理評判有關心理學的言論。就像學生即使把艾瑞克森（Erik Erikson）的人生發展階段論完全忘記了，也仍可運用本書介紹的思維工具，辨別在媒體中不斷湧現的心理學資訊的眞僞。一旦掌握這些終身受用的思維技巧，就有助於我們評判各種理論主張。首先，它將使我們對那些看似合理的事實做初步、總體的判斷。其次，這些技巧可以作爲評估「專家」觀點可信度的標準。在這個複雜的社會裡，人們愈來愈依賴專家的觀點；所以在獲取知識時，判斷專家觀點的可信度就更加重要了。這些批判性思考技巧不只可運用於各個學科或整個知識體系，在心理學領域裡尤其重要，因爲大眾傳媒常常歪曲這個領域的內容。

對任何試圖阻止心理學受到曲解的努力，許多心理學家都抱著悲觀的態度。雖然這不是完全沒有道理，但如果我們就此放棄，這種曲解就會愈來愈嚴重。我這本「消費者指南」式的書，就是想表明，我們不能讓這樣一個惡性循環，成爲纏住心理學的魔咒。

儘管我很高興這本《這才是心理學！》能夠改版這麼多次（編按：本書爲最新的第七版），但有些遺憾的是，本人當初撰寫本書第一版的理由及目的，至今依然沒有改變。媒體對心理學的介紹仍然像以前一樣的誤導大眾，而學生在開始上心理學入門課程時，對心理學的誤解還是與以往一樣多。這導致本書以後幾個版本之出版理由與目的，都和第一版一模一樣。不過令人欣慰的是，我的寫作初衷現在已經得到愈來愈多心理學教師的回響。史丹佛大學（Stanford University）心理學家謝帕德

（Roger Shepard, 1983）表達了與本書寫作初衷相同的看法：「雖然大多數心理學的本科生也許不會走上科學研究之路，但我們仍然希望他們有能力評估那些片面、幼稚、混亂及誇張的報導，這些報導不斷地出現在媒體上，自稱是社會科學的『新發現』……那些認爲可以通過未經驗證的常識，或更糟糕的，可以通過星相學之類毫無實驗根據的僞科學，就能達到對人類行爲和心理現象的充分了解之謬論，勢將不斷地向我們提出挑戰。」（p. 855）

本書的目的是對批判性思考技巧作簡略的介紹，這些技巧將幫助人們更有效的理解心理學的議題，以及他們周遭世界所發生的事。

第七版更新內容

第七版的《這才是心理學！》，結構上沒有很大的改動，因爲第五版已經對其中一個章節做了較大的重組。各章節的內容和順序也保持原樣。應審核者和讀者的要求，這次的版本做了一點「瘦身」。讀者和使用者都不希望本書篇幅增大，因此，實際上，這一版沒有像以前的版本那樣增加篇幅，第七版反而比第六版縮減了15%左右。篇幅雖然縮減，但沒有刪減任何概念，只是刪掉一些冗餘拖遝的例子，保留下最好的（依據審核者和讀者的回饋意見）。

最重要的是，我更新和重寫了書中的許多例子，用最近的研究和議題替換過時的事例。我也極

力引用與本書提及的各種概念和實驗相關的最新研究資料。因此，在這一版中總共有一二九處新引文，讀者在此可以得到所有例子和概念的最新參考資料。

本書出版的宗旨一直保持不變——簡略介紹批判性思考技巧，幫助學生更有效的理解心理學的議題。在一九九〇年代，大學裡強調批判性思考技巧的呼聲愈來愈高（Halpern, 1998）。的確，有些州立大學系統已經進行加強批判性思考教育的課程改革。與此同時，也有其他教育學學者認爲，批判性思考技巧不應該脫離特定的學科內容。而《這才是心理學！》正好融合了這兩個觀點，幫助教師在教授豐富之現代心理學知識的同時，也傳授批判性思考的技巧。

歡迎讀者把對本書的批評意見寄到下面的地址和我交流：Keith E. Stanovich, Department of Human Development and Applied Psychology, University of Toronto, 252 Bloor St. W., Toronto, Ontario, Canada, M5S 1V6. E-mail: KStanovich@oise.utoronto.ca。

第一章　活力十足的心理學

——在科學陣營裡還幹得蠻好的

佛洛伊德盛名之累

如果我們在大街上隨機攔截一百個人，請他們說出任何一個已逝或在世的心理學家，那麼像戴爾（Wayne Dyer）和其他一些曝光率很高的「媒體心理學家」一定會被提及。假若我們把這類心理學家排除在外，讓他們以對心理學做出公認貢獻為標準列出心理學家，那排在首位的一定是佛洛伊德（Sigmund Freud），而史金納（B. E. Skinner）將可能以較大差距屈居第二。至於其他的心理學家，人們很可能再怎麼想也想不出了。由此可見，是佛洛伊德——加上媒體經常出現的通俗心理學家——共同塑造了大眾對心理學的印象。

名聞遐邇的佛洛伊德，不但深遠地主宰了大眾對心理學的認識，也導致大眾對心理學產生許多

誤解。例如，很多上心理學導論課的學生都驚訝地發現，在美國心理學會和美國心理協會這兩個正式的學術組織成員中，把關注佛洛伊德精神分析學派的心理學家加在一起，分別也只占了10%和5%。

事實上，現代心理學研究完全不受佛洛伊德的思想左右，更沒有受限於他的想法，只有媒體和一些人文學科領域才會這麼認爲。在種種現代心理學研究的議題、資料和理論中，他的工作不過滄海一粟。心理學研究和理論包括五位近期諾貝爾獎得主的工作：休伯爾（David Hubel）、卡尼曼（Daniel Kahneman）、西蒙（Herbert Simon）、史培利（Roger Sperry）和維瑟爾（Torsten Wiesel），另外包括美國國家科學基金會（National Science Foundation）的前負責人阿特金森（Richard Atkinson），而這些人都不爲大衆所知。

更不幸的是，心理學與佛洛伊德的過分認同，給這門學科帶來的是誇張和曲解。其中最糟糕的是，佛洛伊德的研究方法已經遠遠不能代表現代心理學研究所運用的手段了（別忘了佛洛伊德開始他那些知名的研究工作是一百多年前的事了）。因此，對佛洛伊德研究方法的理解，會造成對心理學研究的極大誤解。比如，佛洛伊德不做操控實驗，而我們從第六章可以清楚的了解，操控實驗是心理學研究最有力的武器；第二，佛洛伊德透過個案研究證實或證僞理論，而我們可以在第四章中看到這種方法的謬誤之處；最後，佛洛伊德的研究最大的問題，在於他的理論與行爲資料之間的關聯。我們同樣可以在第二章了解到一個理論要達到科學標準，該理論與行爲資料間的關聯必須至少要滿足

幾個條件；而佛洛伊德的理論卻完全不能滿足這些評判準則（Crews, 1996, 1998; MacMillan, 1997; McCullough, 2001; Watters & Ofshe, 1999）。簡言之，佛洛伊德是根據個案和內省資料建立起他的精緻理論，但是這些資料基礎並不牢固。他只是集中精神於建立複雜的理論構架，卻沒有像現代心理學家那樣，確保這些理論構架是建立在可靠的、以及可以被其他研究者重複驗證的行為資料上。因此我們可以說，精通佛洛伊德的那套研究模式，對理解當代心理學的研究方法反而是個阻礙。

在本章中，我們將從兩方面處理這個所謂由佛洛伊德帶來的盛名之累：(1)我們透過闡釋現代心理學的多樣性，能夠清楚地看到佛洛伊德的理論在這一學科中，事實上是處於相對次要的地位（參見 Haggbloom等人，2002; Robins, Gosling, & Craik, 1999, 2000）；(2)我們將探討在心理學領域中，大部分研究共用的研究方法有哪些特點——它們與過時的、佛洛伊德式的知識累積方式是完全不同的——以此澄清大眾對現代心理學研究方法的誤解，從而讓大眾看到他們對佛洛伊德研究的認同，令他們忽略了一個唯一能整合日趨多樣化的現代心理學之特性：用科學的方法理解人們的行為。

現代心理學的多樣化

現代心理學包含各式各樣的內容和層面，這種多樣化多少使整個心理學顯得鬆散或不夠統一。正如美國心理基金會（American Psychological Foundation）傑出教學獎得主葛萊特曼（Henry Gleitman,

1981）對心理學的描繪：「一個鬆散地連結在一起的學術王國，橫跨從生命科學到社會科學這兩個極端的所有領域。」（p. 774）

理解了心理學令人驚歎的廣泛性及調查研究手段的多樣性，對眞正理解這一學科的本質至關重要。從下列的事實中我們可以淸楚看到這一點。美國心理學會包括五十三個分支，每個分支代表一個特定的研究或應用領域（參見表一）。從表中我們可以看到心理學家研究的主題範圍、研究背景範圍和所使用的不同研究角度。同樣，另一個大機構美國心理協會也是分支衆多。一般來講，從表一可以了解到心理學每個分支都是一個專門化的領域。事實上五十三個分支中的每一個，又各自包括相當廣泛的研究領域和子分支，以至於我們很難找到心理學沒有含納的領域。

◎多樣化的意義

學生在學習心理學時，都希望學到一套統一又完整的心理學理論，可以用來解釋所有的人類意識和行爲。有這種想法的學生往往大失所望，因爲他們學到的不是個完整統一的大理論體系，而是許多不同的小理論，各自涵蓋行爲的某個有限層面（Griggs, Proctor, & Bujak-Johnson, 2002; Zechmeister & Zechmeister, 2000）。對於範圍如此龐大的心理學，建立一個大統一理論基礎確實相當困難。實際上，許多心理學家都認爲這種建立大統一理論體系的目標是不可能達到的；儘管如此，還是有些心理學家在某些領域裡，試圖從事整合的工作（Gibson, 1994; Kenrick, 2001; Kimble, 1994; Solso & Massaro,

表一　美國心理學會的分支機構

1. 普通心理學 General Psychology	17. 理論諮詢心理學 Counseling Psychology
2. 心理學教學 Teaching of Psychology	18. 公共服務部門中的心理學家 Psychologists in Public Service
3. 實驗心理學 Experimental Psychology	19. 軍事心理學 Military Psychology
5. 評估、測量和統計 Evaluation, Measurement, and Statistics	20. 成人發展與老化 Adult Development and Aging
6. 神經行爲科學和比較心理學 Behavioral Neuroscience and Comparative Psychology	21. 應用實驗和工程心理學 Applied Experimental and Engineering Psychology
7. 發展心理學 Developmental Psychology	22. 復健心理學 Rehabilitation Psychology
8. 人格和社會心理學 Personality and Social Psychology	23. 消費者心理學 Consumer Psychology
9. 社會議題的心理學研究 Psychological Study of Social Issues	24. 理論和哲學心理學 Theoretical and Philosophical Psychology
10. 心理學和藝術 Psychology and the Arts	25. 行爲實驗分析 Experimental Analysis of Behavior
12. 臨床心理學 Clinical Psychology	26. 心理學史 History of Psychology
13. 應用諮詢心理學 Consulting Psychology	27. 社區心理學 Community Psychology
14. 工業和組織心理學 Industrial and Organizational Psychology	28. 精神藥理學和藥物依賴 Psychopharmacology and Substance Abuse
15. 教育心理學 Educational Psychology	29. 心理治療 Psychotherapy
16. 學校心理學 School Psychology	30. 心理催眠 Psychological Hypnosis

31. 心理學會州支會事務
State Psychological Association Affairs
32. 人本心理學
Humanistic Psychology
33. 智力缺陷和發展性障礙
Mental Retardation and Developmental Disabilities
34. 人口與環境心理學
Population and Environmental Psychology
35. 女性心理學
Psychology of Women
36. 宗教心理學
Psychology of Religion
37. 兒童、青少年和家庭服務
Child, Youth, and Family Services
38. 健康心理學
Health Psychology
39. 心理分析
Psychoanalysis
40. 臨床神經心理學
Clinical Neuropsychology
41. 心理學和法律
Psychology and Law
42. 獨立執業的心理學家
Psychologists in Independent Practice
43. 家庭心理學
Family Psychology
44. 男女同性戀議題的心理學研究
Psychological Study of Lesbian and Gay Issues
45. 少數民族議題的心理學研究
Psychological Study of Ethnic Minority Issues
46. 媒體心理學
Media Psychology
47. 運動心理學
Exercise and Sport Psychology
48. 和平心理學
Peace Psychology
49. 團體心理學和團體心理治療
Group Psychology and Group Psychotherapy
50. 成癮研究
Addictions
51. 男性和男性化的心理學研究
Psychological Study of Men and Masculinity
52. 國際心理學
International Psychology
53. 臨床兒童心理學
Clinical Child Psychology
54. 幼兒心理學
Pediatric Psychology
55. 藥物療法
Pharmacotherapy

注：沒有第4和第11分支

1996）。例如，在過去十年中，由於演化論心理學家的理論貢獻，心理學的內部凝聚性大大增強。這些學者試圖透過將人類的心理過程視爲一系列服務於宗族識別、配偶選擇、合作、社會交換、兒童撫養等對進化功能而言至關重要的關鍵機制，從而實現了這些心理過程可能在概念上統一的願景（Barrett, Dunbar, & Lycett, 2002; Bjorklund & Pellegrini, 2000, 2002; Buss, 1999, 2000, 2001; Cosmides & Tooby, 1992, 2000; Pinker, 1997, 2002）。

但無論心理學家站在什麼立場，基本上都同意建立大統一理論框架是非常困難的，即使眞的有可能，也得很多很多年以後才能實現。缺乏理論的整合，不僅使學習心理學的人失望，也讓許多否認心理學能爲科學做出貢獻的批評者有了把柄。他們的批評往往源自一個錯誤觀念：所有眞正的科學都應該具有一個宏大的、統一的理論根基。這個觀念的錯誤之處在於，它忽視了很多其他現已被稱爲科學的學科，缺乏一個大統一的概念體系。正如哈佛大學（Harvard University）心理學家艾斯蒂斯（William Estes, 1979）明確指出的：

> 實驗心理學家面臨的困境，在整個科學領域並不鮮見。物理學在二十世紀早期就是處於分裂狀態，即使本科教學也是同時講授不同的觀點，以至於我是從機械學、熱力學、光學、聲學和電子學等不同的課程中掌握物理學的。同樣的，化學也曾分為無機化學、有機化學、物理化學和生物化學。當時這些學科之間的交叉融合並不比現在的心理學分支少。

> 物理和化學都是在出現更高抽象水準的數學理論後才實現整合的。與心理學類似，醫學現在也分為很多分支，也還沒有出現大統一的跡象。（pp. 661-662）

一個學科的領域範疇是什麼往往是歷史的偶然，與學科本身沒有太大關係。例如，大學的院系很多時候並不是按照知識本身的結構設置，而是按行政管理上的方便劃分的，因此院系劃分也是可以改變的。正如美國心理學會的前主席彼萬（William Bevan）所言，心理學理論本身四分五裂，而心理學系居然還能存在下去，恐怕更多是出於管理上的方便，或者只是爲了讓系內成員能方便地接收自己的郵件罷了。

想像一所大學廢除心理學系，把它的成員分到其他相關的系並不是一件很困難的事。生理心理學家可以分到生物系，社會心理學家分到社會學系，認知和知覺心理學家分到認知科學系，組織和工業心理學家分到商學院，臨床和諮詢心理學家分到社會工作系、人力資源系或教育系，發展心理學家分到教育系、認知科學系或人力資源系等等。分到各系後，心理學家可能感覺不到和新同事之間在學術交流方面會與舊同事有什麼大的差異，而且有可能覺得與新同事更容易合作共事。由此可見，就內容而言，現代心理學的各領域根本很難整合爲一個彼此相互關聯的整體，所以就需要能夠從更高的層次尋找可以整合心理學這一學科的憑藉。

一旦理解決定學科結構的社會和歷史因素，我們就可以認識到，要求所有領域完全統一整合並

不合乎邏輯。所以**心理學研究**這個術語在某種程度上比**心理學**更能準確反應這一學科的多樣化。這個用法還能讓學生更容易理解，爲何在解釋和預測行爲方面，有一些心理學領域會比另一些領域取得更多及更重大的科學進展。而**心理學**這個詞就表達不出這樣的差異，反而讓人感覺它是一個具有整合性的學科，但這一整合性卻正是這一學科所沒有的。如果我們試圖尋找心理學這一學科的整合性，那一定不要在心理學家研究的課題內容的關聯性上去找；相反地，我們應把重點放在心理學家發展形成新知識的方法。這方面，是我們唯一有可能找到心理學家之間共同點的地方；然而也正是在這方面，是我們發現心理學遭到最多誤解的地方。

在科學上一統

用科學的方法了解人類的行爲，是心理學各分支之間僅有的連貫性。事實上，心理學如果說有任何特殊性，就在於它是廣泛地使用各種科學方法研究人類的行爲。僅說心理學是研究人類行爲的，並不能把心理學與其他學科分開。其他專業團體和學科——如經濟學家、小說家、法律、社會學、歷史、政治學、人類學和文學研究等——也都在某種程度上與人的行爲相關。在這一點上，心理學並非與衆不同。

在實際應用方面，心理學研究也沒有與衆不同。例如，想幫助別人的這個出發點是促使很多大

學生主修心理學的原因，但這也是很多其他學科——如社會工作、教育、護理、職業治療、物理治療、警政學、人力資源和語言治療法（甚至哲學，參見 Marinoff, 1999）等——的一部分。同樣地，透過諮詢幫助他人，也是教育、社會工作、警事工作、護理、神職工作、職業治療等學科的手段。因而，如果只是培養專業人員、透過諮詢的方式幫助別人，並不一定需要開闢一門叫心理學的學科。

其實**只有**兩件事可以讓心理學有成爲一門獨立學科的理由：用科學的技術手段研究人或非人的所有行爲，以及透過科學的方法應用這些研究成果。如果不是這樣，心理學也就失去存在的意義。

心理學有別於其他行爲學科，在於它試圖給大衆兩方面的保證：其一是心理學研究所得的有關行爲的結論是從科學實證中得來的；其二是心理學的實際應用都源於科學方法。心理學在發展過程中是否偏離過這兩個目標？是的，而且經常如此（Dawes, 1994; Fox, 1996; Garry, Frame, & Loftus, 1999; Lilienfeld, 1998, 1999; Loftus & Guyer, 2002; Mook 2001; Watters & Ofshe, 1999）。本書就是想幫助我們掌握達到這兩個目標的途徑。在第十二章，我會再討論心理學家如何因罔顧科學標準而將自己的合法性打了折扣；但原則上，正是這些標準使心理學有資格成爲一個單獨的領域。如果有一天心理學認爲這些目標不值得追求——因而不想再堅守科學的標準——它就可以關門大吉，讓各分支按內容歸併入其他學科，因爲那時它會變成一個完全多餘的學科。

從這一點我們可以清楚地看到，任何人想理解心理學，首先要知道，心理學之所以爲心理學，就是因爲它是一門以實徵資料爲根基的行爲科學。讓人們理解這一點也是貫穿本書的主旨，因爲唯

有認識到這一點，我們才有能力想清楚心理學是什麼。相反地，人們之所以對心理學產生這麼多誤解及困惑，也是因爲他們沒有認識到心理學爲一門科學。例如，我們經常聽到圈外人說心理學不是科學，爲什麼會這樣呢？

讓大衆覺得心理學不可能是一門科學有多方面的原因，我們會在隨後的章節中討論到，人們對心理學到底是什麼的困惑，多半是那些僞心理學提倡者一手精心炮製的。我們的社會已經發展出一個相當大的產業，它們爲自身利益營造出一批僞科學信念系統來迷惑大衆，讓他們認爲所有事物都與心理學有關，並宣稱沒有任何理性的標準可以評價心理學的解說及理論。於是，在這種氛圍影響下，市面上鼓吹類似「催眠瘦身法」、「激發你的潛在心理能量」和「夢中學法語」等概念，營業額高達數百萬美元的心理自助產業就大行其道了。實際上，這些東西不是毫無科學根據，就是與現有證據抵觸。

另一個對科學心理學產生抗拒的力量，是來自不想讓科學進入一些原本是由不容置疑的權威和「一般常識」所支配的領域。歷史上提供許多事實，顯示大衆對科學知識的抗拒；也就是說，大衆更傾向於用哲學的思辯、神學的預言或民間的常識解釋世界。每種科學在發展過程中，都曾經過這一抗拒階段。例如，與伽利略（Galileo）同一時代的其他學者，居然拒絕使用他發明的新式望遠鏡觀測星空，因爲承認有環繞木星運行的衛星存在這一事實，衝擊了他們的哲學和神學信念。幾個世紀以來，人體解剖學的知識發展緩慢，就是因爲受到大衆及教會對解剖屍體之禁忌的影響（基督教

認爲身體的內部是「神的領地」，參見 Grice, 2001）。達爾文（Charles Darwin）之所以不斷受到貶斥，布洛卡（Paul Broca）的人類學會（Society of Anthropology）之所以在法國受到排擠，也都是因爲人類作爲一種生物的知識在當時被認爲對國家具有顛覆作用。

科學的每次重大進步都經歷過巨大的阻力。不過當人們理解到新的發現不但不會詆毀人性，且能因拓展我們的知識而更加成全人性時，反對的聲音也就消失了。現在誰還會認爲星雲圖及天文理論會摧毀我們對宇宙的讚歎呢？誰會選擇屍體解剖開禁前的醫療保健系統而不選擇現有社區內的良好醫療保健呢？所以說，對星際或人體做實徵研究，並沒有詆毀人性。近些年，達爾文的演化論可以說爲遺傳學和生物學的驚人進步打下堅實的基礎。然而，即使是在因爲他的學說而逐漸逼進了解人類本質和起源的今天，殘存的反對聲音依然存在。在美國，極端教派繼續在公立學校鼓吹上帝創世說。研究表明仍有相當比例的大衆（在某些調查中甚至超過半數）不接受人是由天擇演化而來的這一事實。如果連這麼多科學證據的演化生物學都還不斷受到大衆的抵制，那麼，對處於發展階段，目的在對人們長期持有的、有關自己的資訊進行科學檢驗及挑戰的心理學，我們又怎能奢望人們不會置疑呢？

最後要指出的是，不承認心理學是科學的人，對什麼是科學的認識往往也一團糟。在大學裡，幾乎每個心理學教師都會遇到一些選擇主修心理學的一、二年級學生說，他們之所以選擇心理學是「因爲我不喜歡科學」。當教師告知心理學正是一門科學時，他們都感到非常吃驚。教師對他們接

下來的沮喪反應（「我不能相信我還要學統計！」）早已見怪不怪。當教師再問：「你們在入大學之後選修過生物學和化學嗎？」他們的回答一定是：「哦，當然沒有啦！我一直都在迴避理科的課程哪！」這裡，我們看到的是一個極具諷刺性的現象：學生在對科學一無所知的情況下，居然確信心理學不是一門科學！

那麼，什麼是科學？

要明白心理學是什麼，我們必須先明白科學是什麼。不過，要明白科學是什麼，也許得從科學不是什麼談起，以澄清對這個詞的大量誤解。首先，科學不是由研究課題的內容決定的。對宇宙任何方面的研究都可以發展成一門科學，當然也包括對人類行為的研究。因而，我們不可能把宇宙分為「科學」和「非科學」領域。雖然歷史上確實有過一股強大的勢力想把人類排除在科學研究的範圍之外，但是它們都以失敗告終。當前針對心理學是不是科學所提出的懷疑，應該還是這場歷史爭論的餘音。有關這場爭論，我們在下面會提到。

另外，科學也不是標榜非得用實驗儀器不可。科學研究絕非以試管、電腦、電子設備、腦科學技術為特徵（即便用這些指標，心理學也完全符合科學的定義，因為幾乎在所有大學的心理學系都裝滿電腦、化學製劑和各種電子設備）。這些僅僅是科學的附著物，而非本質特徵。科學是一套理解宇宙的思考、

觀察方式及程序，讓我們能進一步地了解它的運作。

在本章的餘下部分，我們將討論科學的三個緊密相連的核心特徵：(1)系統化的實徵主義；(2)公開知識的產生；以及(3)對可解決的問題之探研。雖然以下我們將分別闡述，但請謹記，三者是緊密聯結構成一個整體結構的。（更詳盡的討論請參閱《參考文獻》中Bronowski, Cournaud, Dawkins, Medawar, Popper, Raymo和Sagan的專著。）

◎系統化的實徵主義

用任何一本詞典查**實徵主義**（empiricism）一詞，你會發現它的意思是「基於觀察的做法」（譯注：指用能觀察到的東西爲研究物件的主張），即科學家是透過探究世界來發現世界的。也許這在你看來已是不爭的事實，這正好說明，科學態度已流傳幾百年的歷史了。在此之前，人們對科學的認識並非如此清晰。回想在那個拒絕使用望遠鏡的年代，人們一直以爲知識只有通過冥想或從權威那裡才能得到。當伽利略宣佈在木星周圍觀察到衛星時，另一個學者希茲（Francesco Sizi）試圖用以下言論——而不是觀察——反駁他：

> 人的頭上有七竅：兩個鼻孔、兩隻耳朵、一雙眼睛和一張嘴；天上的星星也有兩顆吉星、兩顆災星、兩顆發光的，還有一顆不置可否的水星。從很多類似的現象，如七種金屬

等，我們可以料定行星的數量也必定是七個……另外，猶太人和其他古老民族，都和現代歐洲人一樣把每個星期分為七天，並用來分別命名七顆行星；現在如果我們增加行星的數量，則整個系統都將瓦解……還有，衛星是肉眼見不到的，也就不會影響地球，既然沒有用，也就沒有存在的必要。（Holton & Roller, 1958, p. 160）

引述這段文字不是爲了說明它的愚蠢可笑，而是想指出這便是曾經作爲反駁實際觀察資料所想出來的、在當時看來相當合理的辯詞！我們對它的輕視和嘲笑都是事後諸葛亮。經過三個世紀實徵方法的洗滌，我們比希兹處於優勢。但如果沒有這些年洗滌的經驗，我們也很可能會頻頻點頭地贊同他。可以說，實徵研究方法從來不是那麼顯而易見的，這就是爲什麼即使在科學占主導的當今社會，我們仍有必要教導人們認識它。

不過僅只是純粹、單一的實徵主義是不夠的，這一點可以從本節標題「**系統化**的實徵主義」中了解。觀察法是好而且必要的，但是單一、非結構的觀察不會發現科學知識。例如，你可以從起床到就寢做整日的觀察記錄，不久你就會累積出大量的事實，但是這並不會加深你對世界的認識。而科學觀察是**系統化**的，因爲它是有結構地觀察，所以其結果能夠揭示世界運行的規律。科學的觀察通常是爲了檢驗理論；它是想透過測試各種不同的說法，理解我們生活的世界。科學的這一結構性可以令我們根據觀察的結果，去支持一些理論，否定另一些。

◎可公開驗證的知識：重驗與同儕審核

從某種特殊意義上講，科學知識具有公開性，但這**公開性**不一定就是要把科學發現貼在社區中心的公告牌上，而是指科學知識不只存在於個別人的頭腦中。一個基本共識是：任何一個科學發現在沒有接受科學界其他人的批評和驗證之前，都不能稱爲科學知識。知識如果是「特別」的——是一個人想出來的，而且不受他人的審視及批評，那它一定不是科學知識。

科學是通過重驗這一機制來實現其可公開驗證性。一項發現如果要在科學界得到公認，必須能讓其他科學家重複相同的實驗，並得到相同的結果。只有這樣，我們才說該發現是被重複驗證了。所以，這種重複驗證機制保證了某個發現不是因爲個別研究者的偏見或錯誤造成的。總之，某發現要得到科學界的接受，前提一定是：可以被最初發現者以外的其他研究者重驗。只有經歷這樣的檢驗，此發現才具有公開性。此時它不再是最初研究者的私有財產，而是可以爲其他研究者繼續發展、批評並應用到自己的領域中。

詩人唐（John Donne）講過「沒有人是一座孤島」，同樣的，在科學上也沒有任何研究者是一座孤島。每個研究者都與科學社團、已有的知識基礎緊密相連，是這種相互連接促進了科學的累積發展。在科學進程中，研究者經常要在先前的知識基礎上從事新的探索和超越，如此，我們就必須保證之前的知識累積方法的確可以作爲進一步研究的基礎。

可公開驗證的知識，是指可以經由科學界其他人重驗、批評和繼續發展的驗證程序得到的知識。這個標準不僅對科學家重要，也是消費大眾評價出現在媒體中的科學資訊的重要依據。第十二章我們會談到，騙子或那些施用僞科學的人，與眞正科學家的一個重要區別在於，前者往往繞過科學界通過科學出版物發表研究發現的常規，直接透過媒體公佈他們的「發現」。在相信那些宣稱是科學的發現之前，一個顚撲不破的評判眞僞方法是先問一問，這些發現是否已經在具有同儕審核制度的正式科學期刊發表了。這個問題的回答，立刻可以區分僞科學和眞理。

同儕審核一般是這樣進行的：投稿文章首先提交給若干科學家審核，審核者將批評意見提交給一名編輯（通常是該刊物涵蓋的領域中，在某方面有多年經驗的科學家），再由編輯決定是否發表該文，或需要進一步的實驗或統計分析，還是因爲存在某個致命問題和缺陷而拒絕採用。一般來說，大部分刊物都會刊登有關該刊物的編審政策及章程，透過它們，我們可以獲悉此刊物是否執行同儕審核制度。

當然，不是所有資訊被同儕審核過的內容都一定正確，但至少它們達到同儕審核的標準。這也只是最低標準，而不是最嚴格的，因爲許多學科領域都有品質各異的刊物。只要達到最基本的標準，大部分的科學想法一般都可以在各種各樣的刊物上公開發表。因此那些辯稱只有範圍極窄的資料和理論才能被科學刊物發表的說法，其實是不正確的。但江湖術士卻常常借此讓媒體和大眾相信他們是被「正統科學圈子」的陰謀所害，堵死了他們出版之路。讓我們看看像心理學這樣的科學領

域，提供了多少合法出路吧，《心理學摘要》（*Psychological Abstractes*）中羅列著上千種學術期刊的文章摘要，而這些刊物幾乎都有同儕審核的制度。因此，幾乎任何稍微經得起考驗的理論和實驗，都可以找到發表的地方。

再次聲明，我並不是說在《心理學摘要》上刊出的研究都是好的或正確的。正如前面所言，它們僅僅是滿足了最低標準。但是我們倒可以說，一個假說、理論、說法或療法，如果不能在學科內部經過同儕審核發表出來，就顯示它有問題。尤其當在證據不足的情況下，又要透過媒體宣傳自己的發現，那麼**恰恰顯示它肯定是偽科學**。

雖然在不同學科間同儕審核的方式不盡相同，但根本邏輯是一樣的。同儕審核和重驗機制是科學保證其客觀性和公開性的根基。思想和實驗只有經過其他人的嚴格審視及尖銳批評，才能開始接受大眾驗證的洗禮。同儕審核機制雖然還未能達到完美，卻是我們消費者僅有的保障機制。任何忽視或輕視它的做法，都會讓科學和偽科學同流合污，並受制於它們的淫威（見第十二章）。在隨後的章節中，我們會更詳細的討論，不按科學心理學程序實踐、忽略審核監督機制時，將給我們帶來多麼巨大的危害和慘痛教訓。

◎能以實徵解決的問題：科學家對可驗證理論的探求

科學是用來解決可解決的、具有明確性的問題；也就是說，科學家解決的是那些有可能用現有

的實徵技術找到答案的問題。如果一個問題不可能有解答，或一個理論目前沒有技術可以驗證，科學家就不會研究。例如，「在照顧三歲以下的小孩時，給予結構化語言刺激的孩子，日後的閱讀學習水準是否比不給刺激的孩子好？」就是一個科學問題，因爲這個問題可以利用現有的實徵方法獲得解答。而「人性本善還是本惡？」和「生命意義是什麼？」則不是可實徵的問題，也就不屬於科學的問題了。

科學的進展通常是透過一連續過程進行的：先提出理論假說對世界的某特定現象進行解釋開始，經過由理論推展出預測、再經實驗驗證，最後到根據實驗結果修正理論（理論〉預測〉實驗〉修正）。所以科學家所說的**可解決的問題**也就是指「**可驗證的**理論」。理論如何才是可驗證的？首要的是理論必須對自然界中的特定可觀察事件有一定的作用；這就是所謂的**可實徵驗證性**。可驗證的標準就是我們會在第二章提到的**可證僞性標準**（falsifiability criterion）。

我們知道，對科學僅只能解決可被實徵驗證的問題這一點，存在兩大誤解。第一，科學家經常被刻畫成僅僅認爲科學問題才是重要問題的形象。當然這不一定是實際情況，僅僅因爲科學家在專業工作中不解決非科學問題，並不能說明他們對其他非科學的工作漠不關心。

第二，我們說科學家只關注可用實徵方式解決的問題，並非意味著所有問題都能分類爲可解決的與不可解決的，也不是說這種劃分是永不可改變的。恰恰相反，現在解決不了的一些問題，隨著理論和實徵技術的發展，有一天會成爲可解決的。例如十年前，關於傑佛遜（Thomas Jefferson）的奴

隸海明斯（Sally Hemings）撫養的孩子是不是傑佛遜的，歷史學家都不相信這是個可用實徵方式解決的問題。可是，一九九八年這問題隨著遺傳技術的進步得到解決。刊登在《自然》雜誌（*Nature*）上的一篇文章（Foster et al., 1998）指出傑佛遜非常有可能是愛斯頓・海明斯・傑佛遜（Eston Hemings Jefferson）的父親。

這個例子也說明了一般科學發展和新科學的產生過程。對什麼是當前可解決或不可解決問題的爭論有很多，科學家自己通常也因這類問題的模糊性而難以達成共識。所以雖然科學家對「只研究可用實徵方法解決的問題」這一標準認識一致，但對要如何具體決定哪些問題是可解決的、哪些是不可解決的卻爭論不休。諾貝爾獎得主梅達瓦（Peter Medawar）就曾把他的一本書定名爲《可以解決的藝術》（*The Art of the Soluble*, 1967），書中他指出，科學的部分創造性，就體現在它是在人類知識的最前線，探索可以用實徵技術進行研究的問題。

心理學裡有很多好例子，說明一個問題可以從不可解決向可解決的方向過渡。很多問題，如「孩子是怎樣學會父母說的語言的？」「我們爲什麼會忘記曾經記得的事情？」「加入一個群體後，人們的思想和行爲是如何改變的？」在有人意識到它們可以通過實徵方法解決這些問題之前，人們早就對這些問題提出自己的揣測。隨著這種可以用實徵方法解決問題之認識的發展，各個領域中有關人們行爲的各項問題，才集中組合構成現在的心理學。心理學的問題，也逐漸從哲學中獨立出來，發展成爲一門實徵的學科。

認知心理學家平克（Steven Pinker, 1997）講過，「無知」可以分爲問題（problems）和謎團（mysteries）兩類。如果是問題，就有可能找到答案，而即使還沒找到答案，我們可以預計未來會找到。對謎團來說，我們就無法預想會找到答案。用平克的這些詞語，我們可以把科學看作把謎團轉變成問題的過程。事實上，平克（1997）在《心智如何運作》（*How the Mind Works*）一書中，就表達了這個意思：「許多有關心智的謎團，不管是關於心像還是戀愛，都在最近由謎團升級爲問題。」（p. 9）

心理學與世俗智慧：「常識」的問題

我們每個人都有一套如何處理自己的人際交往行爲及如何看待自己和他人的內隱理論（implicit theory）。社會、人格和認知心理學家也都研究過個體的這些內隱理論。研究發現，人們很難清晰、邏輯地把這些理論表達出來。事實上，人們只有在刻意關注或發現它們受到挑戰時，才意識到它們的存在。實際上，大多數人並不會遵循一套成型的理論行事；反而，當我們感到要對行爲做出解釋和說明時，往往會引用一些老生常談的諺語或一串的箴言。但是這些關於行爲的「常識」本身往往是相互矛盾的，也因此是不可證僞的（有關可證僞的原則將在下一章討論）。

人們常常用相互矛盾的諺語，在不同時間解釋同類事件。例如，「三思而後行」是一個強調行動要謹慎的好俗語，但是我們也常聽到用「機不可失，時不再來」勸人們行動果斷。「小別勝新婚」

表達了人們對周圍事件的一種情緒反應，「眼不見心不煩」不也一樣？類似的、呈對立狀態的諺語還有「欲速則不達」和「兵貴神速」，「三個臭皮匠賽過一個諸葛亮」和「三個和尚沒水喝」，「安全第一」和「不入虎穴，焉得虎子」，「異性相吸」和「物以類聚」；還有我經常告誡學生「今日事今日畢」，可是當我又和他們說「車到山前必有路」時，眞希望自己沒有說過先前那句話。

這些大受歡迎的陳腔老調，構成我們對行爲的一大堆內隱的「解釋」。無論發生什麼情況，都可以抬出某個詞句解釋一番。難怪我們都自認爲判斷人類行爲或個性的高手。從適用性的角度來說，這些俗語簡直無所不能。心理學家特根（Karl Teigen, 1986）曾讓受試者評價類似上述意義相互矛盾的諺語，結果發現人們確實傾向於同時認爲兩個相互矛盾的諺語都是對的。特根總結出一個有趣的發現：「人們判定某個諺語的正誤，可以和他判斷一個與之對立諺語的正誤完全**無關**。」(p. 47)

因此，有時我們很難駁倒一些內隱的心理理論，但正如下一章會談到的，正是因爲很難駁倒，它們往往也沒什麼大用。更何況，即使其中的某些民間信念是可驗證的，它們往往被心理學研究證實是錯的。例如，對高中生打工一事，大多數美國人都認爲是件好事，理由如下：(1)打工掙到的錢可以貼補他們未來的教育費用和家庭開銷；(2)可以發展工作情操，使他們在日後的職業生涯中能承擔更多的責任；(3)可以幫助他們形成對美國經濟的正確看法；(4)因爲已經融入經濟體系，會提升他們的學習積極性。

發展心理學家曾做了大量有關打工對學生影響的研究，涉及對學生的行爲、態度和學業成績等

層面（Bachman & Schulenberg, 1993; Steinberg, Brown, & Dornbusch, 1996; Steinberg, Fegley, & Dornbusch, 1993）。他們發現，美國文化中對青少年打工的所有世俗看法都是錯誤的。事實上，只有很少的打工收入眞正貼補家庭的開銷或作爲未來的教育支出，大多的收入都花費在比闊的奢侈品或廣告鼓吹的「必需品」上。而且，高中階段的打工經歷普遍對學生不具有教育和提升經驗的意義。更有趣而反諷的是，打工經驗反倒讓學生對工作以及其對經濟所可能帶來的貢獻，變得更反感、更不尊重。在研究中，打工的青少年比沒打過工的，更常抱「拚命工作的人比不拚命的人精神更不正常」和「天下沒有公司眞正關心員工這回事」之類的想法（Greenberger & Steinberg, 1986）。格林伯格（D. Greenberger）和史坦伯格（L. Steinberg）在回顧大量的同類研究後，得出結論：「打工絕對弊大於利，甚至會引發青少年的一些失足行爲。」（p. 6）很顯然，我們給年輕人描繪了一個關於工作的虛幻願景。那些認爲打工可以幫助青少年「塑造人格」和「形成對金錢的觀念」等文化口號，其實都是錯的。因此，我們可以說，這些老生常談就類似人類學家在研究發展中國家時喜歡挖掘的古老傳說一樣，它們令我們感覺舒服些，也令我們當下的文化行爲變得合理，卻毫無現實基礎。

我們還可以輕易的舉出許多實例說明世俗常識或民俗諺語的謬誤。例如，一般大衆認爲在學業上成績出色或經常讀書的小孩，通常在社交及體能方面表現不好。雖然這是完全錯誤的認識，但仍然廣爲流傳。大量事實表明，這類小孩體能優秀，在社交上也更加積極（Gage & Berliner, 1984, pp. 18-19; Market Facts, 1984; Zill & Winglee, 1990）。例如，學業成績良好的孩子，比成績不良的更容易贏得朋

友的接納（Gage & Berliner, 1984），經常讀書的人比不常讀書的更喜歡運動、慢跑、露營、踏青、修理汽車等（Zill & Winglee, 1990）。

許多世俗觀念一旦出現，就長久流行。例如，一九八〇、九〇年代有一種世俗觀念認爲：低自尊直接導致攻擊行爲；但是，實徵研究顯示攻擊行爲和低自尊之間沒有關聯。如果有，反倒是：攻擊行爲和高自尊經常聯結在一起（Baumeister, 1999; Baumeister, Boden, & Smart, 1996; Baumeister, Bushman, & Campbell, 2000）。另一個在九〇年代極其流行的假設是，低自尊導致學業成績不佳。後來證明，自尊與學業成績之間的關係似乎與這一廣爲學校管理者認同的假設正好相反，是優異的學業成績，和其他方面的成就導致高自尊（Kahne, 1996; Ruggiero, 2000; Stout, 2000）。

拉德福（B. Radford, 1999）曾談到一個似乎無法泯滅的世俗迷思：我們僅僅用了我們腦力的10%。這個觀念儘管沒有神經科學的根據（參見 Beyerstein, 1999; Higbee & Clay, 1998），卻流傳幾十年了，而且好像已獲得「心理事實」的地位。拉德福引述專欄作家塞繆森（Robert Samuelson）給「心理事實」下的定義：「雖然沒有堅實的事實證據，卻由於日常生活中的不斷重複，而被認爲是眞實的觀念。」（p. 53）

世俗的信念並非完全可以不靠事實證據而持續。有時，一旦反證廣爲人知，民間信念也會改變。例如，多年前流傳的諺語是「早熟早爛」（Fancher, 1985, p. 141），認爲兒童早熟將意味著日後生活的毀滅。但是現在，由於心理學研究表明這個觀念是錯誤的事實已在大衆文化中生根，我們就再

也聽不到這一謬論了。

這個例子也警示我們小心對待當今的「常識」，因爲要讓昨天認爲正確的常識在今天變爲謬論並不是難事。畢竟，常識是指「人盡皆知」的知識。那麼，就像過去一百五十年來，在美國曾經是每個人都知道的「常識」，如「女人不應該有投票權」、「美國黑人不應該接受教育」或「殘障人士都應該送進收容所」等，現在不都被我們否定爲垃圾了嗎？主要就是因爲它們都沒有實徵研究的支持。從這些例子我們可以看到，相對於常識而言，心理學的用武之處，在於它以實驗爲基礎，驗證這些常識背後的假設。有時，這些常識的假設並不能通過檢驗，從而使充當檢驗者角色的心理學難免與文化信念發生衝突。心理學也因而成了勇敢的揭發者，讓這些早爲人們慣性接受的世俗智慧再無立足之地。難怪很多人不但不信心理學的檢證報告，也對做出這些報告的心理學家視若無睹。

作爲一門新興科學的心理學

在實徵心理學的發展過程中，反對的聲音總是不絕於耳。一百年前，英國劍橋大學（Cambridge University）拒絕建立心理物理實驗室，因爲這類研究會將「人的靈魂放在量尺上，從而侮辱了宗教的神聖。」（Hearst, 1979, p.7）心理學家米爾（Paul Meehl, 1991）提過：「在當代社會科學使用的實驗和統計方法尚未發展時，對人們只靠『人盡皆知』所獲取的知識（這裡的人是指只要是人類的一員即

可），幾乎從來不曾進行過任何批判性檢驗。」（p. 440）基於此，心理學欲建構可實證之課題的戰爭可謂剛剛取得勝利。但是隨著科學的進展，許多問題逐漸變成可以用實徵方法解決，心理學家將以此挑戰人們更多的固有觀念。諸如，道德理性的發展、戀愛心理學、種族偏見的本質，和宗教信仰的心理及社會決定因素等，都是心理學家正在探索的問題。最近，關於幼童性行爲的研究引發很大的爭議（Hagen, 2001; Rind, Tromovitch, & Bauserman, 2001）。有些人仍然反對就這一領域進行實徵研究（Hunt, 1999），但心理學在這一領域還是取得相當的進展。

列文和歐唐納（J. R. Levin and A. M. O'Donnell, 2000）曾論及，某些人反對心理學研究的原因在於他們宣稱自己「不需要知道」。他們敘述，當一個校董會提出要家長決定是否讓孩子在有不同年齡學生的班級學習時，有教師提出應該先做研究，但學校董事會否決了這一建議，因爲他們認爲不管「多齡」教學或「單齡」教學哪種方法有效，家長都會強迫學校完全轉向較好的那種教學方式。列文和歐唐納（2000）從而提到：「（對於可能的研究結果）校董會根本就不想知道！」（p. 66）因此，我們應該充分意識到人們常常否定心理學研究，並不是因爲他們覺得不好，而是想逃避由此可能帶來的麻煩。

也許你已經意識到，對心理學的批評中有個很大的矛盾。一方面，一些人反對把心理學稱爲科學，否認心理學家可以經由實徵驗證有關行爲的事實；另一方面，另一些人由於害怕心理學家可能發現某些行爲領域的事實，而反對心理學家在這些領域裡做研究。史金納行爲學派的心理學家總是

必須面對這類矛盾的批評。例如，反對者說行爲主義的強化規律不適用於人類行爲。同時，另一些反對者又害怕人們會運用這一規律規範及控制他人的行爲。行爲主義者因此腹背受敵，一些批評者否認行爲主義者找到的行爲規律是可以應用的，另一些批評者則害怕這些規律被濫用。

以上這些現象的產生，主要由於心理學是一門新興的科學，剛剛開始探索以前人們沒有想到要研究的行爲問題，它的青澀多少也解釋了爲什麼它自身作爲一門實徵科學的地位一直受人質疑。在第十二章我們也將討論到，心理學自身發展過程中的所作所爲，也對外界的這種誤解產生推波助瀾的作用（Dawda & Martin, 2001）。這些行爲忽略了心理學是一個依靠實徵研究方法的學科，從而模糊了心理學與其他領域的重要區別。由於心理學起步較晚，許多研究領域現在還處於嬰兒期，剛開始步入研究成果收成期，以及建立有解釋力之理論的階段。但無論如何，過去的四十年裡，心理學已經穩固地建立在我們稱之爲科學、其中各學科緊密聯結的知識結構中。

摘要

心理學是一個主題非常廣泛、但相對鬆散的學科。它包括一些不能用同一內容概括的衆多研究主題。然而它可以透過「都是用科學方法研究行爲」這一標準整合爲一門學科。科學方法絕不是指一套生硬的規則，而是指一些非常通用的原則。最重要的三點是：(1)科學採用系統實徵主義的研究

方法；(2)它是以可公開驗證的知識爲研究目標；(3)它努力尋求用實徵方法可以解決而且可以發展出能付諸驗證之理論的研究課題（下一章的主要內容）。系統實徵主義所指的有結構的及有控制的觀察，是本書隨後幾個章節的主題。科學通過同儕審核的程序和重複驗證的機制，以保證知識的公開性。

心理學是一門新興科學，因而經常會和世俗智慧衝突。這種衝突是任何新興科學都會遇到的，了解這個衝突，有助於我們理解那些對心理學作爲一門學科的敵意。但是與世俗常識之間的衝撞，同時也令心理學成爲一個非常讓人振奮的學科。現在很多人加入這個領域，正是因爲它可以提供機會來檢驗那些存在了幾百年、被廣泛認可的「常識」。

第二章 可證偽性

——如何阻擋腦袋中的鬼靈精到處搗蛋

一七九三年，費城爆發一種嚴重的傳染病——黃熱病。當時費城的一位名醫、也是美國《獨立宣言》的簽署者之一拉許（Benjamin Rush），據說是少數幾個可以治癒黃熱病的內科醫師。拉許的藥物理論認爲，凡伴有高熱症狀的疾病，都必須透過大量放血（用手術或水蛭吸走體內的血液）醫治。他用這種方法治癒很多病患，包括自己。然而，有人卻指責他的放血療法比黃熱病本身更恐怖。黃熱病流行過後，拉許更加堅信放血療法的療效，雖然事實上已經有好幾個接受放血的病患在治療中死去。他是怎麼得到這一結論的？

有人這樣描述拉許的觀點：「一方面堅信自己的藥物理論是正確的，另一方面卻缺乏有效的方法對治療效果進行系統研究以證實自己的觀點。因此，他把每個病情好轉的病例都歸結爲治療方法的功效，而把每個失敗的病例歸結爲病情本身的嚴重性。」（Eisenberg, 1977, p. 1106）換句話說，如果

病患情況好轉，就被視為放血療法有效的證據；如果病患去世了，則表明病情已經嚴重到**任何**治療方法都無法奏效的地步。現在我們知道對拉許的這種批評是正確的：他的放血療法的確和黃熱病同樣危險。本章，我們將討論拉許究竟錯在哪裡。他的錯誤在於違反了科學思維的一個重要原則，此一原則同時也是評價心理學理論特別有用的一項標準。

本章還要詳細探討第一章提到的一個話題，關於科學的第三個普遍特徵：科學家只能研究可解決的問題。科學家所說的**可解決的問題**通常意指「可驗證的理論」，他們通過保證理論在實徵觀察時，有可以驗證是錯的可能性（即「可證偽性」），以確保其可驗證性。也就是說，一個理論所欲解釋的行為現象，一定要在現實生活中能表現出來，以致可通過觀察事實或現象是否存在，得知該理論對不對。我們將在下面看到為什麼**可證偽性標準**在心理學中這麼重要。

理論與可證偽性標準

拉許在評估放血療法的療效時掉進一個致命的陷阱。他評價證據的方法使他不可能得出該療法無效的結論，因為如果病患的康復是對療法的肯定（從而也是對他的醫療理論的肯定），那麼病患死亡就應該意味著是對其療法的否定，這才合理。但是，他卻把這種否定合理化了。拉許這種解釋邏輯違反了建構和檢驗科學理論的一條重要原則：他讓他的理論永遠不可能被驗證是錯的。

在表述科學理論時，必須遵循一個原則：從理論推導出的各種預測要有證僞的可能性。在評價某一理論的新證據時，必須看在蒐集新證據時，是否有可能證實該理論是錯的，這個原則就是**可證僞性標準**。著作在科學工作者中間廣爲流傳的科學哲學家波普（Karl Popper），就一直竭力強調可證僞性標準在科學進程中的重要作用（Magee, 1985）。

可證僞性標準是：要說一個理論是有用的，那麼根據它所做的預測必須是具體的；也就是說，理論在告訴我們如果它是對的則哪些事情應該發生的同時，還必須說清楚哪些**不會發生**。一旦後者**眞的**發生了，我們就清楚地知道：這個理論有些地方出錯了。這時就有必要對該理論進行修改，或提出一個全新的理論取代之。這兩種做法都會使我們提出的理論更接近眞實情況。相反地，如果一個理論的預測包括所有可能觀察到的資料，這個理論將永遠不需要也不會修改，我們也就將停滯於現有的思維窠臼而不會取得任何進步。因此，一個成功的理論，並不是那種可以解釋或預測任何可能出現的事實或現象的理論，因爲這樣反而恰恰喪失了該理論的預測能力。

由於在本書餘下的章節中，我們經常會對理論進行評估，在這裡有必要先澄清這個經常被誤解的名詞：**理論**。人們掛在嘴邊的一句口頭禪，就是「噢！這僅僅是個理論罷了」。從這句話不難聽出，**外行人**在使用**理論**一詞時，意思是指一個未證實的假設，僅只是一種猜測，一種預感。但這**絕不是**科學意義上的**理論**。科學家的理論，不是指沒有證實的假設。

科學的理論是指一套具有內在關聯的概念，這套概念可以解釋一組既有的資料（所反映的事實及

現象），並且能對尚未進行的實驗結果進行預測。**假設**則是指由理論（相對假設而言具有較概括和較廣泛的意義）推衍出來的更為具體的預測。現存比較好的理論都是因為由其推衍出來的假設大都得到了驗證；也就是說，這些理論的結構到目前為止是與衆多的實徵觀察結果吻合的。但是，一旦觀察資料開始與理論推衍出來的假設產生矛盾，科學家就會重新開始建構另一個新的理論，以便給觀察到的資料提供更好的解釋。因此目前正在進行科學討論的理論，都是在某種程度上被實徵驗證過、所做的預測到目前為止還沒有和觀察到的資料產生矛盾的理論，因此不僅僅只是猜測或預感而已。

外行人和科學家使用**理論**這一術語的差異，經常被一些虔誠的正統基督教徒利用，作為將上帝創世說帶入公立學校講堂的支持證據（Pennock, 1999）。他們的辯詞通常是「畢竟主張人類是由物種演化而來的演化論僅僅是一個理論而已。」這種論調中所用的**理論**，就是**外行人**理解的「僅僅是一種猜測」的意思。然而，主張天擇的演化論並不是外行人所說的理論（而應是他們理解的**事實**），而是**科學**意義上的理論，它是有大量觀察資料支持的概念體系（Berra, 1990; Maynard Smith, 1998; Ridley, 1996, 1999）。它不等同於那種純屬猜測的說法。著名生理學家多布詹斯基（Theodosius Dobzhansky, 1973）寫過一篇很有名的文章〈生理學中除了演化論別無他物〉，說明這一理論的重要性。

◎敲門節奏理論

下面讓我們舉個例子說明可證偽性原則如何作用。我的一位同事認為不同類型的人敲門節奏會

愈具体 愈精確 愈可能被推翻

不同，他說他可以據此推測來訪者為何許人物。例如，某天一位學生敲我辦公室的門時，在我開門之前，這位同事預測敲門者是女性。我打開門，果真是一名女學生。我對該同事說，他的理論很難說服我，因為即使不根據他的「敲門節奏理論」推衍，他也有50％的機率猜對。同事聽了之後說他的理論可以有大於50％的預測率。又來了一個人敲門，這次該同事告訴我是二十二歲以下的男性。我打開門，是一位我認識的男學生，他剛剛高中畢業，小於二十二歲。於是我告訴該同事，這一次他的理論有點說服力了，因為我們學校二十二歲以上的學生很多，以下的卻很少，因此如果只靠瞎矇，猜錯的機率較高。不過我仍然堅持這次還是太容易猜，因為年輕的男性在校園裡本來就很常見。同事見我很難侍候，於是提出再做一次檢驗。當第三個人敲門時，他預測：「女性，三十歲，五呎二吋高，左手拿著一本書和一個錢包，右手敲門。」我打開門，發現他的預測完全正確。這一次，我的反應截然不同了。如果這三個人不是與我同事預先串通好的，現在他的理論可以說完全說服了我。

為什麼我會有如此不同的反應呢？從一開始的「那又怎樣？」到後來的「哇！」答案就在預測的具體性和精確性有所不同。愈具體的預測在得到證實後，愈能給人留下深刻的印象。但是請注意，預測的具體性是直接與其證偽性掛鉤的。愈具體、愈精確的預測，就愈可能被將來觀察到的資料推翻。比如，有太多的人並**不是**三十歲、五呎二吋高的女性。我對同事之預測的不同反應，隱含著這樣的意思：一個理論的預測愈能夠把依其理論認為**不**可能發生的事情之範圍擴大，該理論就愈

能給我留下深刻印象。

好的理論所做的預測，總是顯示自己可以被證僞。差的理論則不會置自身於被證僞的危險中，它們所做的預測通常非常概括性，以至於總是驗證爲正確的（比如，下一個敲門的人年齡在一百歲以下），或是用完全不可證僞的語句表述（像拉許的例子：「治癒就是理論正確，不癒就是病患已經病入膏肓。」）。事實上，一個理論如果置於不可證僞的保護下，它便不再是科學。這也正說明爲什麼波普是在試圖劃分科學與非科學的差別時，提出可證僞性原則的重要性。這一原則與心理學關係密切，也與我們在第一章中有關佛洛伊德的討論直接相關。

◎佛洛伊德與可證僞性

上世紀初的幾十年裡，波普一直在尋找一個疑問背後的答案：爲什麼有些理論使科學知識停滯不前，而另一些卻導致知識的長足進步（Hacohen, 2000）？例如愛因斯坦（Albert Einstein）的相對論，它之所以能夠導致後面一系列驚人的發現（例如，從遠距離恆星發出的光線在行經太陽附近時會彎曲），正由於這一理論的預測結構與很多有可能發生的事件或現象矛盾，因此其理論是很有可能被證僞。

波普指出，一些使知識停滯的理論卻並非如此。例如佛洛伊德的精神分析理論，就是用複雜的概念結構對人類行爲進行事後解釋，而不做事前預測。波普認爲，正是因爲它可以解釋一切已發生的事情，所以在科學上毫無用處——因爲它不做具體的預測。精神分析學派的追隨者，花費大量時

間和心血使該理論能夠解釋所有已知的人類事件，從個人的怪癖到廣泛的社會現象；但是他們在讓該理論成功地成爲事後解釋的厚實依據時，卻使其喪失了任何科學價值。與其把佛洛伊德的精神分析看作當代心理學的理論，不如當作文學創作的靈感更爲恰當（Robins & Craik, 1994; Robins, Gosling, & Craik, 1999, 2000）。它在心理學領域的消亡，部分原因就是不能滿足可證僞性這一標準。（譯注：最近有大衆媒體〔*Newsweek*, 2002/11/11, pp. 46-47〕報導，佛洛伊德理論已經在認知神經科學領域，用腦造影技術得到支持。這可以說是該理論多年受正統科學心理學排斥以來，一個「揚眉吐氣」的突破。這也說明科學方法及技術的日新月異，眼前沒法得到支持的理論，並不一定就是錯的。但是不管科學的方法及技術如何進步，科學驗證的邏輯及程序是不變的。）

正如一位知名評論家指出的：「不正確但廣泛傳播的有關心理學的想法，無可避免地會對整個社會帶來災難。由於精神分析學派的聲望一度被人爲地抬高，導致社會上許多有疾病及遺傳缺陷的人延誤有效的治療，轉而從個體早期經歷中尋找自己現有痛苦的根源。」（Crews, 1993, p. 65）再以妥瑞氏症（Gilles de la Tourette syndrome）的治療史爲例，這種病的症狀是發病時軀體的某個部位會抽搐和痙攣，並伴有言語症狀，例如發出咕噥聲和狗吠聲、模仿言語（echolalia，無意識地重複他人的語句）、穢語癖（coprolalia，強迫性重複淫穢語句）等。這是一種器官型的中樞神經系統失調症，現在已經可以使用藥物氟哌啶醇（haloperidol）根治（Bower, 1990, 1996a）。然而在它的治療發展史上，妥瑞氏症患者卻飽受種種殘酷的迫害。早期他們被宗教權威視爲是妖魔而遭迫害，後來又要被迫接受「軀

邪」的儀式（Hines, 1988）。更重要的是，在一九二一至五五年間，精神分析學派的概念佔據了對該病症之病因及治療研究的主導地位，許多學者用無法被證偽的精神分析解釋理解該病，而這些似是而非的解釋，模糊了人們對該症本質的認識，從而阻礙對眞正病因及治療的研究（參見 Kushner, 1999）。例如，一位學者曾經這樣說：

> （妥瑞氏症是）精神分析學派理論對腦疾病研究產生消極影響的一個典型例子。最初認為這種病是腦的退化過程引起的；但本世紀初期，當佛洛伊德的理論成為時尚，人們的注意力開始偏離大腦……結果是使該類疾病的病患轉而求救於使用精神分析法的精神科醫師，而不是神經科大夫，從而忽略了對身體器官的檢查和研究。（Thornton, 1986, p. 210）

夏皮羅（A. Shapiro）等人（1978）曾提到當時許多精神分析師對妥瑞氏症的一些現在看來很可笑的想法。其中一位認爲他的病患「不願意放棄抽搐，是因爲對她而言，這種抽搐已經變成性快感的源泉和潛意識性慾的表達方式」；另一位精神分析師則認爲這種抽搐「已經等同於手淫……與生殖器相聯結的慾力（libido）錯放在身體的其他部位」；第三個把這種抽搐看作是「肛門－受虐轉化期的症狀」；第四個認爲妥瑞氏症患者具有「強迫性人格和自戀傾向」，病患的抽搐「代表一種情感的症狀，是對想要表達情感的壓抑性防禦」。夏皮羅等人（1978）下面的一段總結，恰當地說明了忽視可證偽性這一標準可能對理論帶來多壞的影響：

> 精神分析學派建構理論的方式實際上面面俱到。抽搐是一種轉化期症狀，它表現為歇斯底里但又反映了性慾，它受意志控制而又是強迫性的，它是器官性病變，同時又與心理衝突有關……所有這些心理學的標籤、診斷和治療，很不幸地統統貼到病患及其家人身上。而且在這樣做的過程中，治療者自大、武斷，從而給病患及其家屬造成很大的傷害……由於這些論文後來產生廣泛的影響，因而給人們認識和治療這個疾病造成極大的障礙。（pp. 39-42, 50, 63）

後來，當研究人員認識到，精神分析理論的「解釋」對治療該疾病毫無用處時，進一步認識和治療妥瑞氏症才變得可能。但那些毫無用處的解釋往往相當誘人，因為它們看似在解釋事情——事實上它們確是解釋了所有事情，不過都是事後諸葛。精神分析理論的解釋，只創造了理解的假象而已。由於要努力解釋所有事後發生的事，所以理論本身得不到進展。一個理論只有在不預測**任何事件**都可能發生，而只是具體地預測，即事前告訴我們什麼具體的行為會出現時，進步才會發生。當然，由這類理論推導出的預測有可能證實是錯誤的，但這正是科學研究的長處，而非缺點。

◎會隱形的小鬼靈精

如果人們能夠從研究的問題中抽離出來，就不難察覺哪些理論的構想是不可證偽的；特別是如

果能汲取前人錯誤的教訓（就像拉許的例子），就更容易察覺了。另外，當事例很明顯是編造的，我們也很容易察覺出其不具可證僞性。例如，我發現人的大腦中有一個可以控制行爲的潛在機制，一開始這一發現還鮮爲人知。但不久之後，你就會在一些隨處可購得的八卦報刊上讀到有關這一發現的報導。這一潛在機制是這樣的：在大腦的左半球靠近言語區域，住著兩個小鬼靈精，他們有能力控制發生在大腦多個區域的化學放電過程。也就是說，這兩個小鬼靈精基本上是控制了一切。不過他們很狡猾，有本事覺察一切對大腦的探測活動（無論是透過手術或是X光之類的途徑），一旦他們感覺到外界的探測，就會悄悄躲藏起來（哦，我剛才忘記說，他們有隱形的本事），所以人們永遠無法證實他們的確存在。

你也許認爲我用這樣一個騙小學生的故事做例子，是侮辱了你的智慧。但是如果你能設想，你像我一樣作爲一個心理學導論的授課老師，卻經常被學生問及爲什麼我在課堂上不和大家講有關超感官知覺（extrasensory perception, ESP）這一領域的驚人發現，或是爲什麼不講過去幾年在通靈學取得的進展，也許你就不會怪我爲什麼會用這麼幼稚的例子說明問題了。回答這些學生提出的這些幼稚的問題，我不得不告訴他們：他們聽到的大部分有關這些方面資訊，無疑都來自大衆媒體，而不是有學術價值的科學刊物。事實上，有些科學家已經在調查及重驗這些媒體炒作的研究成果，但是都不成功。我進而提醒這些還沒有完成心理學研究方法課程的學生，一個研究成果能否成爲科學事實，並被人們接受，是取決於它的可重驗性，特別是在我們遇到研究結果與過去蒐集到的資料或現

有理論預期的矛盾時，這一點尤爲重要。

我進一步指出，很多科學家對有關超感官知覺的研究已經到忍無可忍的地步，原因之一固然是該領域已經被詐騙、江湖郎中、媒體剝削搞臭了。更重要的卻是如迦納（Martin Gardner, 1972）指出的，科學家察覺到超感官知覺研究的不可證僞性。

超感官知覺的研究通常是這樣進行的：首先，「相信者」（即在研究開始前就相信超感官知覺存在的人）宣稱已經在實驗室裡證明超感官知覺的存在。然後，一位「懷疑者」（對超感官知覺的存在持懷疑態度的人）受邀到實驗室證實超感官知覺的存在。通常懷疑者在檢查完實驗的內外環境之後，都會建議相信者對實驗條件進行更嚴密的控制（關於實驗控制的類型，我們將在第六章詳細討論），以減少相信者從中做假的可能性。雖然有時這些建議會遭到拒絕，但是大部分善意的相信者會「從善如流」。一旦這些較嚴謹的控制條件都設置好，超感官知覺就不復存在了（參見 Alcock, 1990; Druckman & Swets, 1988; Hooft, 2000; Humphrey, 1996; Hyman, 1992, 1996; Marks, 2001; Milton & Wiseman, 1999; Wiseman, Beloff, & Morris, 1996）。

然後，懷疑者就會說：實驗的失敗說明，原來的實驗是因爲缺乏有效的控制，才會得出超感官知覺存在這一錯誤的結論。但他們常常會驚訝地發現相信者並不這樣認爲。相反地，這些相信者會用不可證僞的方式解釋超感官知覺：他們堅持認爲心靈的力量——例如超感官知覺——是非常隱晦、細緻，而且很容易因干擾而失靈的。懷疑者「負面的顫波」很可能是這一心靈力量瓦解的罪魁

禍首。一旦移開懷疑者反對的氣氛，這種心靈的力量肯定會再回來。

令人驚訝的是，聽完我解說相信者的心態之後，學生不但沒有失笑，反而恍然大悟的說：「哎呀！這很對呀，我說嘛，原來是懷疑者發出的負面干擾把超感官知覺給搞沒了。」一聽到學生的反應，我雖稍感驚訝，但仍然保持鎮定，並耐著性子提醒他們想想在幾個星期前的課堂上，我舉的會隱形的小鬼靈精的例子。我說，當時全班同學都認爲小鬼靈精的存在證據因爲一要探研它就隱身不見了是件非常荒謬的事，現在相信者提出的超感官知覺的存在證據，與會隱形的小鬼靈精具有同樣的性質。超感官知覺運作的方式與小鬼靈精一樣——只要你不闖進去仔細觀察它，它就好好的在那裡；但是，你若闖進去想看個究竟，它馬上隱形消失。如果我們接受這種解釋，就永遠也無法證明給懷疑者看到它的存在，因爲依相信者的解釋，任何懷疑者的出現都是令其失靈的原因，它存在的證據只能在相信者面前顯現。顯然，這種說法是科學不能接受的。在科學的領域裡，我們沒有磁力學家和非磁力學家（亦即那些相信磁力存在和不相信磁力存在的物理學家）之分。這一點學生顯得有點膽怯，他們咕噥著說同意我的觀點後，迅速離開我的辦公室。

當然，我永遠不知道學生的這種同意是意味著眞正理解，還是尷尬因此想儘快逃離現場。也許兩者兼而有之，因爲對學生來說，他們正面臨一件世界上最困難的情境：面對與自己堅信的正好相反的事實。嚴謹地運用科學準則驗證自己長期固守的信念是相當不舒服的，但這是一個心理學家必須做的。從來沒有人說這很容易。

◎並非所有的驗證都一樣

可證偽性原則對我們如何看待理論驗證具有重大意義及影響。很多人認爲一個好的科學理論是已反覆驗證的理論，他們認爲驗證的次數對評價一個理論來說非常關鍵。但是之前理論的可證偽性的討論，讓我們看到理論的驗證次數並不是關鍵。正如前面的「敲門節奏理論」例子，並非所有的驗證都是一樣的。驗證能否令人信服，取決於理論推衍出來的預測本身可否證偽。驗證二十個實際上不可能證偽的預測（例如小於一百歲的人），遠遠不如驗證一個非常具體、有可能證偽的預測（例如，女性，三十歲，五呎二吋高，左手拿著書和錢包，右手敲門），更能令人信服。

看來，我們不應只注重理論驗證的次數，還應該注重驗證本身的品質。採用可證偽性作爲評價標準，可以幫助運用研究成果的人抵制偽科學聲稱的全能（可以解釋一切的）理論的誘惑。這些全能理論不可避免地會妨礙我們對世界和人類本質的進一步認識。無可否認地，這些理論死巷有時相當具吸引力，正是因爲它們無法證偽。它們在現代這個動蕩混亂的世界裡，常被視爲穩定的避風港。

波普經常談到：「這些（不可證偽）理論之所以會在心理上對人們產生這麼大的吸引力，奧祕就在於它們能夠解釋一切。知道無論發生什麼事，你事先都可以找到解釋它的答案，不僅讓你覺得你掌握了知識，更重要的，它還給你應對這個世界所需的安全感。」（Magge, 1985, p. 43）然而，獲得這種安全感並不是科學研究的目的，因爲它以知識發展的停滯作爲代價。科學是一個不斷挑戰先

前存有觀念的機制，透過實徵研究設法證明之前理論的缺陷。此一特點常常使科學——尤其是心理學——與所謂的世俗智慧或一般常識發生衝突。

◎可證僞性與世俗智慧

心理學可以說是威脅到人們運用第一章提到之世俗智慧的安逸感。因爲作爲一門科學，心理學不能安於用這種駁不倒的解釋來理解人的行爲。心理學就是要對各種不同有關行爲的理論，逐一進行檢驗與篩選。當然，心理學很歡迎世俗智慧中表述清晰、可站出來接受檢驗的內容，而且很多這類內容也已經整合到心理學理論中了。只不過心理學不會尋求由事後可以解釋一切、但事前做不出任何預測的解釋系統所提供的安逸感。心理學也不接受那些永不需更新、代代相傳的世俗智慧系統。如果你我不向學生或大衆坦承這一事實，我們就像是用石頭砸自己的腳，自求毀滅。遺憾的是，有些心理學教師或主張將心理學大衆化的人卻往往擔心，心理學對世俗智慧構成的這種威脅會困擾人們，於是就傳遞一些錯誤的訊息以舒緩這些困擾：「你將學到一些有趣的知識，不過不用擔心，心理學不會挑戰你深信不疑的東西。」這種說法是錯誤的，因爲它會令人們對什麼是科學，以及什麼是心理學更加疑惑。

科學**尋求**的是想法的改變。科學家是在努力描繪世界的眞實面目，而不是描繪我們固有的信念所認爲我們應該有的世界之面貌。在現代思潮中有一股危險的趨勢認爲：我們要儘量避免讓一般大

衆知道事實的眞相——在這些人面前遮上一層無知的面紗是必要的，因爲他們不曾學會如何處理眞相。這樣的觀點，心理學和其他科學一樣都會予以駁斥。生物學家吉瑟林（Michael Ghiselin, 1989）曾進一步指出，如果知識得不到傳播，我們全都會蒙受損失：

> 如果鄰居都很健康，我們也會更好。為了比別人更健康而壟斷所有的藥物供應，不給周圍的人用，是非常愚蠢的行徑。對待知識也一樣。鄰居的無知就如他們的身體不健康一樣，對我們自己也毫無益處，而且還可能是造成我們不好的原因。工業生產及我們所有人都受益於熟練的勞動力。我們都要依賴其他人的技能和特長。（p. 192）

當我們周圍充斥著對人類行爲有誤解的人時，心理學家也會像吉瑟林那樣，感到我們都蒙受損失。世界是由大衆對教育、犯罪、健康、工業生產、兒童福利，以及其他關鍵問題的衆多態度塑造而成的。如果這些態度都來自不正確的行爲理論，所有人都會受到傷害。

◎承認錯誤的自由

可證僞性原則最有用、最具解放性意義的地方，在於它不認爲在科學上犯錯是一種罪過。被證僞了的假設也能夠爲科學家提供資訊，令他們調整自己的理論，從而更符合觀察得來的資料。哲學家丹尼特（Daniel Dennett, 1955）說過，科學的本質在於「在大衆面前犯錯」（p. 380）。當蒐集的資料

與理論不一致時，透過不斷修改理論，科學家最終能建構出一個反映世界本質的理論。可證偽性是理論內在的更新機制，我們不應像處理日常事務那樣迴避這些錯誤。

事實上，如果在處理個人日常生活事務時也使用可證偽性原則，也許我們的生活品質會大有改善。這就是前面我使用**具解放性意義**一詞的原因。解放性是特別針對它對個人的意義而言的，因爲由可證偽性原則引發的意義已經超越科學意義。如果我們明白，當深信的事情與觀察到的事實相悖時，我們最好是改變信念而不是否認事實，將使這個世界減少很多社會和個人的問題。物理學家奧本海默（Robert Oppenheimer）說過：

> 在任何人的推理思維過程中，總有個時候他會回顧自己的過去說：「這個問題我當時沒想明白。」對某些人而言，他們的職業要求他們要有毫不羞愧地承認「我的看法不對」的能力。科學研究是讓這個認錯的過程每時每刻都可能發生。往往你會在思考很長時間的一個問題上，發現矛盾及不尋常的現象。這些現象帶來的震驚，可能促使你重新思考。與有些自認見識廣博、不斷試圖尋找一些合理說法為自己最初犯下的錯誤辯解的人相比，可謂大相逕庭。（Dos Passos, 1964, pp. 150-151）

請試想一下，過去有多少次，當你與他人激烈爭論時，突然間——也許正當你剛剛給予對方一次猛烈的反擊以維護自己的觀點時——發現在某些關鍵問題或關鍵證據方面，你的看法的確有錯。

當時你是怎樣處理這個錯誤的？你可曾向對方承認錯誤，並且肯定對方的看法現在看來的確比你的觀點更合理？你或許不曾這樣做。如果你和大多數人一樣，一定會「沒完沒了地尋找一些勉強說得過去的說法，爲自己的錯誤辯解」。你會拚命使自己在不宣告失敗的情況下從爭論中擺脫出來。承認自己錯了恐怕是你是絕對不會做的，但其實它卻是你唯一能做的。因爲如果辯論雙方在爭論之後，都仍然不知道到底誰的觀點更接近眞理，如果爭論的內容沒有公之於衆（不像科學研究），如果雙方爭論的勁道勢均力敵，如果對爭論的結果（例如知道自己錯了）不能正確的回饋給對方（像前述的例子），那麼我們將找不到適當的機制使人們的信念與事實眞相吻合。這就是爲什麼有這麼多私下或公開的對話及溝通令人更加困惑，也是爲什麼心理學在解釋人類行爲的原因方面，比起一般常識或世俗智慧，是更可靠的資料來源。

很多科學家都見證了，在科學的進程中，犯錯是正常的事，並且知道科學發展的眞正危險在於人類總想避免暴露自己的錯誤觀點，明白這兩點是相當重要的。諾貝爾獎得主梅達瓦（1979）曾寫道：

> 雖然提出有瑕疵的假設情有可原，遲早它都會被比較好的假設換掉，但這樣仍會對曾經深信不疑的人造成嚴重傷害，因為那些熱愛自己理論假設的科學家，大多不願意承認檢驗結果證實自己是錯的。有時候，他們會不讓自己的理論假設接受嚴格的檢驗，而是繞過

這類檢驗，僅僅讓假設中無關痛癢的部分接受檢驗，不冒可能被徹底推翻的危險……對任何一個時代的科學家，我所能給的最好建議是：**一個假設被堅信爲眞的程度，與它是否確實爲眞，完全是兩回事。**（p. 39；黑體字原文爲斜體）

許多著名的心理學家都採納了梅達瓦的建議。在一篇報導著名實驗心理學家克勞德（Robert Crowder）職業生涯的文章中，引述了他一位同事本納加（Mahzarin Banaji）的一段話：「他是我所知道的、最不維護自己理論的科學家。如果你有辦法指出他的理論有漏洞，或他的實驗結果有局限和缺陷，他會非常樂意和你一起計畫如何推翻該理論。」（Azar, 1999, p. 18）艾澤（B. Azar, 1999）描述了克勞德如何發展出一個有關記憶的成分的理論：「分類前的聽覺記憶體」，然後仔細地設計實驗研究來證僞自己的理論。另一個例子是演化論心理學家圖比（John Tooby, 2002），在講述達爾文在科學史上做出里程碑式的貢獻是得益於他對科學的一些態度時，他做過以下這樣的精彩評論：「達爾文能比同時代的人走得更遠，是因爲他比較不爲要整個宇宙來配合自己的理論預測所羈絆。」（p. 12）

但是要讓**科學**發揮作用，並不需要每個科學家都具備隨時要證僞的態度。布朗諾斯基（Jacob Bronowski, 1973, 1977）在他的很多著作中提到：科學揭示世界眞知的獨特力量，並不取決於科學家具備的、超乎常人的某些獨特素質（例如，他們是完全客觀的，他們在解釋結果時永遠不會發生偏差等），實際上，這種力量源於有一個可以糾正會犯錯的科學家之理論的審察過程。在這個過程中，一些科學

家會竭力挖掘和批評另一些科學家的錯誤。哲學家丹尼特（2000）也表示同樣看法，他認爲沒有必要要求每位科學家表現出如克勞德那樣的客觀性。像布朗諾斯基一樣，丹尼特強調「科學家和其他人有同樣的缺點、同樣會犯錯，不過與一般人不同的是，他們意識到其錯誤是源於自己及其所屬學術團體，因此會設計一套精細的系統約束自己，減少犯錯，努力防止因自身的弱點及偏見而影響到自己的研究結果。」（p. 42）

本質上，我們在這裡談的是學術的誠信——以及一種任何科學家都珍惜的開放性。用一般大眾的話來說，開放性是指爲了追求眞理，對一個現象的所有可能解釋都持容忍及開放的態度。但是對科學而言，開放性思維不僅意味著開放的態度，它還有更深一層的含義。哲學家阿德勒（Jonathan Adler, 1998）這樣教導我們：科學更看重開放性思維的另一面，即「一個眞正具有開放性思維的人，是那些願意跟著證據走的人。這種人尊重客觀公正的研究成果而不是受制於個人的偏見……科學的研究方法是跟隨世界的眞相調整自己，而不是要與我們個人的想法保持一致性。」（p. 44）

◎徒有想法不值錢

從前面驗證世俗智慧的討論中，我們得到另一個有關可證偽性原則的摘要推論：想法是不值錢的。當然，這裡的意思是，有些想法是不值錢的。生物學家暨科普讀物著名作者古爾德（Stephen J. Gould, 1987）對這個觀點做了如下解釋：

寫了十五年的每月專欄，我收到大量非專業人士的信件，涉及科學的林林總總……在這些信件中我發現一個出現頻率遠遠高於別的、最常見的錯誤觀念。人們在信中會告訴我他們發現一個革命性理論，一個可以拓展科學界限的理論。這些理論，通常要幾頁寫得密密麻麻的紙才能陳述完畢，它們全是對我們最根本及最深奧的終極問題的一些猜測——生活的本質是什麼？宇宙是怎麼來的？時間的起點？然而這些想法都是不值錢的，因為任何稍有智力的人都可以在吃早餐前想出幾個念頭來。科學家也有能力就這些終極問題提出各種想法，但是我們不這樣做（或者可以反過來說，我們儘量避免思想中出現這類空泛的念頭），因為我們無法想出驗證這些想法的方法及程序，從而判斷它們是否正確。一個看似吸引人卻無法證實或證偽的觀點，對科學來說又有什麼用？（p. 18）

對古爾德最後一個問題的回答是：「一點用也沒有。」他說的那些不值錢的想法，就是前面波普提到的那些宏大理論。這種宏大理論是那麼廣泛，那麼「模糊」，以至於我們可以用它來解釋一切。建構這種理論，主要是給人們提供情感支持，因為它們的提出並不是準備給他人修改或拋棄的。古爾德告訴我們，這樣的理論對科學的進展毫無用處，縱然會讓人感到很安慰。科學是一個創造的過程，然而，這種創造只用於建構符合實徵資料的理論之上。這是很困難的。那些用以解釋世界眞實面目的想法，就不能說是不值錢的了。可能這正說明爲什麼好的科學理論少之又少，而不可

證僞的僞科學信念體系卻遍佈各處——比起前者，後者是太容易建構了。

科學理論與世界聯繫緊密，它們是可證僞的，它們所做的預測是具體的。沒錯，從事科學研究是艱苦的。無可否認的，建構一個眞正提供科學解釋的理論是一項艱辛的工程，但是認識科學運作的一般邏輯卻**並不**難。事實上，現在已有許多專門寫給兒童看的、關於科學邏輯思維的刊物問世（Kramer, 1987）。因此科學對**科學家**來說雖然可能是困難的，從某個角度看，它對一般大衆而言卻應該不是難事，只要大衆願意學習一些普遍原則——不會多於本書介紹的這幾個原則。關鍵是人們必須認眞地掌握這些原則。好，下面就讓我們開始學習吧。

科學中的錯誤：一步步接近眞理

爲了解釋清楚可證僞性原則，我們曾先勾勒出一個科學進步的簡單模型：理論提出，依據理論推衍出假設，然後使用一系列的實驗手段（我們將在之後的章節中討論）驗證假設。若假設被實驗證實，則理論得到一定程度的鞏固；若證僞，那麼理論必須做某種修改，或被更好的理論取代。

當然，我們說科學的知識都是暫時成立的，由理論推衍出的假設有可能是錯誤的，並不意味著什麼人都可以插一腳，什麼想法都可以成爲理論。科學裡很多事物間的關係被多次證實屬實之後，已經成爲定理，因爲它們幾乎不可能被將來的實驗推翻。比如說，人體血液不再循環或地球不再繞

太陽旋轉的情況，將來不大可能出現。這些眾所周知的事實不屬於我們所談的假設，它們不是科學家的興趣所在，因爲它們已經充分證實了。科學家只對自然界那些已有一定的知識，但還有許多未解之謎的問題感興趣。也就是說，他們對那些還沒有證實到已經沒有什麼疑點的問題感興趣。

科學實踐中，科學家只看重未解決的問題而忽視被充分證實的觀點（所謂的定理）這一態度，一般大眾頗不能理解。科學家似乎總在強調他們不知道的東西而不是強調已經知道的。實際情況的確如此，不過科學家這樣做是有充分理由的。爲了獲得知識，科學家必須站在已知事實的前線，當然，前線正是不確定事物最多的地方。科學的進展也正是在這些地方，透過其不確定性遞減的努力過程中產生的。這就經常令科學家在大眾面前顯得非常「不確定」，但這只是表面現象。事實上，科學家只有在知識的前線才感到不確定，而這個知識的前線地帶，正是科學即將有突破的地方。科學家**不會**對那些很多研究證實的事實不確定。

這裡還必須強調，當科學家談到用實際觀察證偽一個理論，或用新的理論代替已證偽的舊理論時，他們並不會完全拋棄該理論原有的、已有實徵資料支持的部分（我們將在第八章用大篇幅討論這個問題）。恰恰相反，新理論必須能夠解釋舊理論能解釋的和不能解釋的所有事實。因此，理論的可證偽性，並不意味著科學家非得建構一個全新的理論不可。科普讀物作家艾西莫夫（Isaac Asimov）在他〈錯誤的相對性〉（The Relativity of Wrong, 1989）中準確地闡釋理論的修正過程，文中記述我們如何逐步認清地球形狀的這一例子。首先，他提醒我們不要認爲原來把地球看作是扁平的觀點是很愚

蠢的，因爲在大平原上（大部分的文明社會與文字發展都起源於大平原），地球確實看起來很扁平。緊接著艾西莫夫提醒我們去想像，如果對不同的理論進行定量比較，結果會告訴我們什麼？首先我們可以根據每哩顯示多少曲率來表述不同的理論。認爲地球是平面的理論會說曲率是每哩0度。現在我們當然知道這個理論是錯誤的。但在某種角度看，它很接近眞理。正如艾西莫夫（1989）所說：

大約在亞里士多德（Aristotle）逝世一個世紀後，希臘哲學家埃拉托斯特尼（Eratosthenes）注意到，太陽光在不同緯度投射的陰影長度不同（假設地球是平面，所有的陰影都應該一樣長）。根據陰影的不同長度，他計算了地球球面的大小，結果算出地球圓周為二萬五千哩，那麼這個球體的曲率就達到每公尺 0.000126度。你可以看到，0.000126與 0的差別有多麼細微。這一細微的差異，讓我們花費好長的時間才從把地球看成平面轉到發現它是球體。這裡要提醒你，即使是像 0和 0.000126之間這樣細微的差異，也是極其重要的。這個差異說明了很多關鍵問題。如果不考慮這一細微差異，如果不把地球看作球體而視為平面，我們就根本不可能準確地描繪出地球上大塊的土地。（pp. 39-40）

當然，科學並不滿足於得到地球是圓的這樣的知識。正如前面討論的，科學家總是盡可能地提煉自己的理論，並且挑戰現有知識的界限。例如，牛頓力學理論預測地球不是一個標準的球體，這一預測後來證實了。現在已經證明地球是一個赤道凸出、兩極稍扁的球體，有時候也把它叫做**扁球**

體。從北極到南極地球的直徑爲七、九〇〇哩，赤道的直徑爲七、九二七哩。地球的曲率也因此不是恆定的（不像正圓球體，其曲率各處都一樣），它在不同的地方每哩會有7.973~8.027吋不等的差距。正如艾西莫夫（1989）所說的，「比起從平面到球體，從標準球體到扁球體的修正要小得多。因此，雖然認爲地球是標準球體的觀點並不正確，但是嚴格地說，比起認爲地球是平面的觀點而言，它沒有錯得那麼厲害。」（p. 41）艾西莫夫所述關於地球形狀的例子，向我們展示科學家是在什麼樣的情況下，使用**誤解**（mistake）、**錯誤**（error）、**虛假**（falsified）這類的術語。使用這樣的術語，並不意味著被檢驗的理論每方面都是錯誤的，而是說它還不完善。所以當科學家強調知識是暫時成立的、可能被將來的研究結果修正時，意思就像這裡舉的例子一樣。科學家相信地球是個球體的同時，他們已經認識到，仔細來看，這一概論有可能做進一步的修改。然而，從標準球體變到扁球體，並沒有完全改變地球是個球體的正確性。所以當我們某一天睡覺醒來時，絕不會發現地球變成立方體。

摘要

科學家通常所說的一個**可解決的問題**，是指一個**可驗證的理論**。可驗證的理論在科學上的定義非常特別：是指可能證偽的理論。如果一個理論不可證偽，那麼它對眞實世界毫無影響，也因此就

毫無用處了。心理學界充斥著種種不可證僞的理論，令它深受其害，這也是造成心理學發展遲滯的原因之一。

好的理論是指那些做出具體預測的理論，也因此它們具有很高的可證僞性。愈是具體的預測，一旦得到證實，給理論提供的支持就愈大。簡而言之，可證僞性原則的一個重要內涵就是：並非所有理論的驗證及證實都一樣有價值。可證僞性愈高、預測愈具體的理論證實之後，就愈能被人們接受及重視。即使預測不被證實（即證僞），對理論的發展也是有用的。一個證僞的預測顯示其理論依據必須拋棄，或須依據實徵資料加以修正。如此一來，被證僞的預測還是有用的，因爲它促使理論做出調整。正是透過這種調整，像心理學這樣的科學才能更接近眞理。

第三章　操作論與本質論

——「不過，博士，這到底是什麼意思？」

物理學家眞的知道地心引力是什麼嗎？我的意思是**眞的**知道。**地心引力**一詞眞正的**含義**是什麼？其內在本質又是什麼？在提到地心引力一詞時我們究竟是指什麼？如果刨根問底，它到底是怎麼一回事？

諸如此類的問題，反映了人們對科學的一種看法。哲學家波普稱之爲本質論（essentialism）。本質論認爲，能稱之爲好的科學理論，必須根據現象背後的本質或內在屬性對現象做出終極性解釋。持本質論觀點的人，一般會認爲無法對現象做出終極性解釋的理論都是無用的，不能對現象做出終極性解釋，就無法反映內在眞實情況——世界的本來面目。

本章將討論爲什麼科學並不回答這類本質論者提出的問題，而是借助於發展出概念的**操作性定義**獲得進展。

爲什麼科學家不是本質論者

事實上，眞正的科學家從不認爲自己是在追求本質論者孜孜以求的那類知識。對本章一開始提出的那些問題的適當回答是：就探求本質的意義而言，物理學家的確不知道地心引力是什麼。科學家也不宣稱要求取完美無瑕的知識（科學的獨特力量並非在於它是一個與錯誤絕緣的過程，而在於它提供一個途徑以消除存在於我們知識體系中的錯誤）。科學也從不試圖回答關於宇宙的「終極」問題。梅達瓦（1984）寫道：

> 科學所不能回答的，以及在科學進展的預見範圍內也不可能回答的問題，總是有的。這些問題是孩子經常會問的——「終極性問題」……我所想到的這類問題包括：一切事物從哪裡開始的？我們都來這個世界幹什麼？人幹嘛活著？（p. 66）
>
> （但是）即使科學不能回答這類關於最初及最終的問題，並不意味著它必須無條件接受對另外一些問題的答案；也不能理所當然的認為既然能提出這類終極性問題，就一定可能找到這類問題的答案。以我們目前的理解力而言，這類問題是無從回答的。（p. 60）
>
> （最後）別忘了，就科學能夠解決的問題而言，它的能力則是不可限量的，除了人們道德方面的缺陷——例如缺乏勇氣——沒什麼能阻礙科學知識的發展。（p. 86）

因此，如果某個人、某個理論或思想體系聲稱能就那些終極性問題給出絕對答案，科學家就會對之產生懷疑，原因之一在於：科學家認爲這類「終極性」問題是科學研究無法回答的。更進一步說，斷言知識的絕對性和完美性，將扼殺對知識的不斷探索。由於能對知識進行自由而廣泛的探究是科學研究的必備先決條件，科學家因此一貫質疑那些聲稱已經發現終極答案的說法。

◎咬文嚼字的本質論者

具本質論傾向的人通常有個共同特點：在還未開始研究術語和概念所涉及知識前，就過分地關注如何定義它們。他們常用的口號是：「不過，我們必須先把要用的術語定義好。」「這個理論概念眞正的**含義**到底是什麼？」似乎意味著：在一個詞可以用做某個理論中的概念以前，必須對使用這個詞所涉及的各種語言問題有個全面而且毫不含糊的認識。事實上，這恰與科學家的工作方式相悖離。在研究物理世界之前，物理學家決不會陷入「如何使用**能量**一詞」的爭論中；或者當我們談到物質的最基本組成成分時，科學家也不會陷入「**粒子**一詞是否眞正表達了我們所指的含義」的爭論中。

只有在廣泛研究與某個術語相關的現象**之後**——而不是之前——才能確定相應科學概念的含義。對概念性術語之定義的修正，源自在科學研究過程中累積的、具有內部一致性的資料和其理論基礎之間的相互修正，而不是對語詞用法的爭論。本質論往往將我們帶入無盡的語詞爭論中，許多

科學家都認爲，玩味這些詞語遊戲會分散我們對更重要問題的關注。例如，針對「生命一詞的眞義是什麼？」這一問題，兩個生物學家的回答是：「**生命**這兩個字在我們的學術領域中沒有什麼眞正的意義可言，它只是一個能夠滿足生物學家研究目的的術語，不是詞義爭論的焦點。」（Medawar & Medawar, 1983, pp. 66-67）。簡言之，科學家的目的是解釋現象而不是分析語言。到目前爲止，所有科學發展的關鍵都在於拋棄本質論而採納操作論（operationism），這在心理學中尤爲明顯。這正是本章探討的主題。

◎操作論者連結了概念與可觀測的事件

好，既然科學概念的意義不是得自對語詞的討論，那麼它們從何而來？正確地使用一個科學概念的標準又從何而來？要回答這些問題，我們必須先討論什麼是操作論，因爲它對在科學研究中如何建構理論，以及對如何評估心理學中提出的理論都至關重要。

儘管操作論有多種形式，對科學資訊的使用者來說，以最廣泛的角度理解它就可以了。**操作論**就是這樣的觀點：科學理論的概念，必須以某種形式植根於或連結於可以通過測量觀察得到之事件。將理論概念與可觀測事件連結，就是這一概念的操作性定義，同時也使這個概念公開化。這種操作性定義排除了概念中可能涉及的、有關個別研究者情感與直覺的因素，並使任何一個執行這個可測量之操作的人都能對此定義進行驗證。

舉個例子，把**饑餓**定義爲「我胃裡陣陣難當的痛苦感覺」就不是饑餓的操作性定義，因爲它與「陣陣難當的痛苦」此一個人感覺經驗連結，而這個經驗是不能被別的觀察者觀測到的。相反地，如果定義像「未曾進食的時間」或像「血糖水準」這樣可以觀測的生理指標，就是一個操作性定義。因爲此一定義是任何人都能取得、可觀察到的測量。同樣地，心理學家不能僅將**焦慮**定義爲「有時會感到的不舒服，緊張的情緒」，而是要將該概念用一系列操作——例如用問卷調查或生理測量等——定義之。前面那個非操作性定義依賴人們對身體狀況的個別解釋，因而無法被其他人重複；而操作性定義則把概念在科學領域公開化了。

認識到科學概念，是由一組的操作而不僅由一個單一行爲事件或任務來定義的，這一點非常重要。包括在這組中的操作，各自都在觀測同一個概念，但是手段略有不同（我們將在第八章中詳談與此有關的統合操作）。例如，教育心理學家用在某一標準化測量工具上的表現來定義**閱讀能力**這一概念，伍德庫克閱讀精通量表（Woodcock Reading Mastery Tests）（Woodcock, 1987）就是常用的工具。該量表總的閱讀分數，是由幾個子測驗中所得的分數相加而來的。這幾個子測驗分別測量與閱讀有關但又有細微差別的技能。例如，先讓受試者閱讀一篇文章，然後進行語詞填空（找個適當的語詞填入文章的空格中），舉出一個詞的同義詞，正確地讀出一個較難的單詞，以及其他一些測驗。在這些測驗上的總體表現，就界定了**閱讀能力**這一概念。

操作性定義逼使我們想清楚如何用外界可觀測的事件或任務界定一個概念。現在，請試想我們

要給一個看來相當簡單的概念——打字能力——下一個操作性定義，因爲你想比較兩種教人打字的方法。請試想一下，爲完成這一比較你必須做的各種選擇及決定。你當然必須測量打字的速度，但要學生打多長的一段文章呢？一百個字的文章顯然太短，一萬字當然又過長。那到底多長才算好？打字速度要維持多久才能與研究者理論建構中的**打字能力**這一概念相匹配？要學生打哪一類的文章才好？應不應該包括數字及數學公式，或不常見的行間？如何處理打錯字的問題？在測量打字能力時，好像應該將速度及錯誤率同時考慮進去才對。但要同時納入這兩個指標，又該如何計出一個總分？是給這兩個指標同等的權重，還是其中一個比另一個更重要？尋求一個好的操作性定義，會讓你不得不審慎思考上述這些問題，這樣也會讓你對打字能力這一概念本身有個更徹底的探討。

不同概念與其對應的可觀測操作性定義之間的關聯，有的比較直接，有的比較間接，程度上有很大差異。很少有科學概念是完全用由現實生活中直接可觀測的操作來定義的，大部分研究概念的定義都是用與之有間接關聯的操作。例如，某些概念是用一組可觀測的操作和與這個個別概念有理論相關的其他概念來定義的。最後，有些概念不是由任何可觀測的操作直接定義，而是用與其他可觀測概念的關聯來定義的。這類概念通常稱爲隱性概念，它們在心理學的研究中屢見不鮮。

舉例來說，由於A型（Type A）行爲模式與冠心病的發病率有關（Austin & Deary, 2002; Curtis & O'Keefe, 2002；Dembroski & Costa, 1987），所以這方面有大量的研究，我們將在第八章中詳細討論。在此僅想著重闡明，事實上，A型行爲模式是由一組子概念定義出來的：具有強烈的競爭意識、潛在

攻擊性、做事急躁、較強的成就動機等等（Dembroski & Costa, 1987）。然而每個定義A型行爲模式特徵的子概念**本身**，也是一個需要進行操作性定義的概念。事實上，許多研究者也確實在操作性定義每個子概念上下了相當大的功夫。對我們現在的討論較爲重要的一點是：A型行爲模式是個複雜的概念。它並非由操作直接定義；相反的，它是由一組其他概念所定義，而這些概念則有操作性定義。A型行爲模式就是一個間接操作性定義的典型例子。儘管理論概念與其可觀測的操作之間關聯的緊密程度不同，但所有概念在一定程度上，都需要透過這些與之相關的實際觀測來獲得其意義。

◎科學概念的演化

科學概念的定義並非一成不變，而是隨著對其觀測的不斷累積及豐富而逐漸演變。如果一個概念原始的操作性定義在理論上證明是無效的，就要換用另一組操作來定義它。因此隨著相關知識的不斷累積，科學概念也不斷的發展並逐步趨向抽象化。例如，有一陣子曾認爲電子是繞著原子核旋轉、帶負電荷的小粒子。但現在，在特定實驗條件下，它被視爲具有波浪屬性的或然率密度函數（probability density function）。

在心理學中，智力這一概念的發展就是類似的例子。最開始，這個概念只有一個嚴格的操作性定義：智力是由智力測驗測到的東西。隨著研究資料的不斷累積，將智力與學生成績、學習、腦損傷與其他行爲，以及生物因素等連結在一起，智力這一概念也不斷的豐富及擴展（Deary, 2000, 2001;

Ferrari & Sternberg, 1998; Sterberg, 2000; Sternberg & Kaufman, 1998; Sternberg & Wagner, 1994）。現在看來，最合適的智力概念，是由幾個更具體的資訊加工操作所定義的概念組成的一個抽象概念。這些假設存在的資訊加工過程，則可通過測量某些行爲表現來直接賦予操作性定義。

記憶理論的概念也歷經類似的演化。心理學家現在很少使用諸如**記憶**和**遺忘**這樣寬泛的術語，而是去測量幾個更具體、代表不同記憶過程的記憶特徵。例如：短期聽覺記憶、圖像記憶、語義記憶及情節記憶。最初記憶和遺忘的概念，已被細化爲具體的操作性概念了。

因此，理論術語的使用是在科學實踐活動中、而非咬文嚼字中發展出來的，這是對科學持操作性態度的操作論及追求純粹定義的本質論之間的最顯著差別。神經學家戈施文（Norman Geschwind, 1985）對這種差異有如下的描述：「我認爲，從醫學的發展史中能了解一件事：許多人認爲研究問題就是要先定義這個問題然後再研究。但這種方法屢屢證明是錯誤的，因爲正確定義問題的唯一途徑，就是先知道答案。」（p. 15）

哲學家丘奇蘭德（Paul Churchland, 1988）也強調：科學概念的含義是從觀察或與其有關的其他概念中獲得的，而不是依靠對詞語的定義。

要徹底理解「電磁場」這個概念，我們必須熟悉它身處的理論原則體系。只有根據對整個體系的認識，才能了解電磁場究竟是什麼以及它有什麼作用。這個例子是很典型的。

理論術語的含義，一般而言，是不能只從單一的、明確的、闡述了所有其被應用的充分必要條件的定義中獲得，而往往是由它身處的理論原則體系推導出其定義。（p. 56）

科學概念在逐步的發展過程中，會與不同的理論體系錯縱複雜地交織在一起，因而獲得不同的操作性定義。這種情形並不代表概念本身有什麼錯。可惜，許多人不相信心理學，就是因爲它的許多重要理論概念，如智力等，擁有多個不同的操作性定義（Sternberg, 2000）。事實上，這種情況並非心理學獨有，也沒什麼大了不，在科學的領域中很常見。例如，「熱」這個術語就可以分別從熱力學和動力學這兩套理論體系做不同的構想及定義，物理學也並沒有因此而遭到貶斥。再如電子，它的許多特性是從視之爲一種波的角度加以理解的，而另一些特性則又必須從視其爲粒子的角度來考慮，才更能理解。到目前爲止，還沒有一個人會因爲物理學存在著這些可相互替代的構想，就提議要捨棄它。

心理學中的操作性定義

提到物理、化學，許多人能夠理解操作論的必要性。科學家在準備研究某一特殊化學反應或討論能量或磁場現象時，人們能夠理解科學家必須用一些方法測量這些要研究的東西。但遺憾的是，

一旦把這些步驟放到心理學的研究上，他們往往就糊塗了。要知道，心理學術語也是需要有直接或間接的操作性定義，才能建立一些有價值、並具解釋力的科學理論概念。為什麼這一點人們總是認識不清呢？

原因之一是心理學所帶有的**預設偏見的問題**（我們在第一章中提過）。人們在研究地質學時，不會帶著具有情感色彩的信念去探看岩石的性質。然而他們在看心理學的研究時，情況則大不相同。每個人心中似乎都各自早有一套有關人格和人類行為的理論，因為人們在日常生活中，始終不斷地解釋著自己及他人的行為。而且，每個人的這些個體心理學理論，也都含有諸如**聰明**、**侵略性**、**焦慮**等理論性的概念，因此人們自然就會問：我已經知道這些概念的意思了，為什麼還要接受另一些定義呢？儘管表面上看來這種質疑似乎很有道理，但其實這對任何欲理解人類行為的科學發展來說，絕對是個阻礙，也是促使大眾對心理學迷惑不解的原因。

大眾對心理學產生誤解，以及媒體在準確介紹心理學發現時存在的種種障礙，主要是因為心理學的許多專業概念及術語，都是以日常生活用語表述的。這些日常用語讓相當廣泛的錯誤觀念得以先入為主地進入人們的腦海中。因此，沒有受過心理學訓練的外行人很少會意識到，當心理學家在使用諸如**智力**、**焦慮**、**侵略性**以及**依附**等詞語作為理論概念時，它們與日常慣用、相同名稱的詞語的意思並不一定相同。

在先前對操作論的討論中，我們已經明顯看出心理學術語與一般大眾所用的術語之間這種差異

的本質。諸如**智力**或**焦慮**等專業術語，必須依靠其直接或間接的操作性定義，決定它們是否在一個心理學理論中被正確地使用。而這些術語常常是極爲技術性及特殊性的，並且在許多方面都不同於日常生活中的用法。例如，許多人在聽到「對大樣本的認知任務進行因素分析所得到的第一主成分因素」這樣一句話時，不會意識到這是**智力**這一心理學術語之操作性定義的一部分。

同樣地，**憂鬱**一詞在外行人的詞典中，就意味著「某種陷入低谷的感覺」。相形之下，《心理疾病診斷統計手冊》（*Diagnostic and Statistical Manual of Mental Disorders*, American Psychiatric Association, 1994），對重鬱症的技術性定義的描述，共占了約十二頁的篇幅，並且與「某種陷入低谷的感覺」根本是兩回事。臨床心理學家定義的憂鬱，和外行人認爲的可以說完全不同（Hollon, Thase, & Markowitz, 2002）。其他學科其實也存在同樣的問題，只不過不如心理學這麼嚴重。回想一下先前對**生命**這一概念的討論吧。正如梅達瓦與梅達瓦（Medawar & Medawar, 1983）指出的：「麻煩就在於，像科學中許多其他專業術語一樣，『生命』一詞是從人們日常生活的用語中借過來的。但是用在科學情境中，其含義已大大不同於一般談話中的用法了。」（p. 66）

當心理學家和外行人用同一個詞語指涉不同的事物時，就很容易彼此產生誤解。如果我們可以創造出一些新名詞表示心理學家研究的概念，這種混淆就會少很多。其實，我們時不時也確實造出一些這樣的新名詞。就如物理學家有他們的**爾格**和**焦耳**一樣，心理學家也有自己的**失調**（dissonance）和**編碼**（encoding）。實際上，這些名詞並不完全是新創的，只是因爲在日常用語中比較生僻，反而

因此能夠避免這裡所說的混淆。

外行人也許會反對我們發明新鮮名詞：「爲什麼心理學家要給我們製造這麼多麻煩呢？用這麼多的行話、生僻用語和高度專業化的定義！我們幹嘛要它們？爲什麼我心中認爲的『智力』就得不到他們的認可呢？」

這裡我們看到一個誤解心理學研究的典型例子。這種誤解在媒體對心理學研究的報導中經常見到。有一份全國性的報紙以「你能用一般人聽得懂的話重新再說一遍嗎？」爲題，報導了美國心理學會一九九六年年會的情況（Immen, 1996）。這篇文章指出：「心理學家用的完全是只屬於他們自己的語言。」該文嘲諷了在會上報告的一份標題爲〈用 Gf-Gc 理論解釋結合 WJ-R 及 KAIT 兩測驗所做的因素分析結果〉的論文。雖然這個記者表示「我甚至無法猜測這個題目到底是什麼意思」，但是幾乎所有受過嚴格訓練的心理學家都知道，這個題目指的是一篇有關智力測量理論方面之新進展的報告。不過，我們沒理由責怪這名記者做這樣的報導，我們不應奢望他聽說過 Gf-Gc 這個名詞，因爲它是心理學研究智力理論的一個技術上的新進展。就如同我們不會期望該記者知道有關一個基本粒子物理學家剛發現、尚未命名及經過議論的中微子的詳細資料一樣。然而，不知怎麼搞的，記者對科學術語的無知（這是完全可以理解的），往往成爲人們用以藐視現代心理學的依據。

現在，讓我們看看解決這個問題的關鍵在哪裡。第一步是要強調先前討論中提到的一點：操作論不是心理學研究獨有的，而是所有科學研究都要求有的。大多數情況下，我們很容易接受並認識

到這一先決條件是必要的。如果某位科學家發現一個放射現象，我們理所當然地認爲他用來測量放射的方法，必須是其他研究者也可以用之觀察到，並能以此獲得同樣研究結果的。操作性定義因此導致科學的公開化，而公開化則是科學的關鍵特徵。只要兩位科學家就同一個操作性定義達成一致，就可以相互檢驗對方的結果了。然而，這個在別處看起來很明顯的道理，一到心理學，有時就變得模糊了。人們經常意識不到對如**智力**、**焦慮**這樣的心理學概念進行操作性定義的必要性。前面說過，原因是我們天天都在使用這些術語，難道我們還不「知道」它們是什麼意思嗎？

對這一問題的答案應該是：「是的！我們的確不知道。」這句話中的「我們」當然不是指科學家，而是一般大衆。一個科學家則必須先「淸楚」他研究的智力的含義，並且考慮如何精確地定義及測量，才能讓其他實驗室的研究者也可以進行同樣的測量，從而才有可能看看在針對此概念所做的研究中，是否可以得出相同的結論。這樣的操作性定義對具體性和精確性的要求甚高，因此這些科學術語與日常談話中，人們爲達到即時的相互理解所用的、意義較含混的口語用詞大相徑庭。

◎操作論推動人性化

過於依賴我們所謂「早都知道了，何必再給出新的定義」這一心態所引發的問題之根源，同樣也是我們依賴「直覺」（即非經驗）信念體系所會引發的問題之根源。你對某事的「了解」可能與張三李四有所不同，那我們如何判斷誰是對的？你可能會說：「可是，我對此感覺非常強烈，強烈到

我**知道**我是對的。」但是，如果張三與你的看法有些不同，而他對他的看法的感覺比你還強烈，那該怎麼辦？然後，如果李四與你和張三兩人的想法又都不同，而他宣稱他一定是對的，因爲他覺得自己的感覺比張三**強烈得多**。

這個簡單而荒謬的例子，只想說明科學知識的一個基本特性：知識之所以爲正確，並不在於它是否爲個別人士對該知識所持有的感覺及相信程度。這個基本特性，事實上是推動人類歷史實現人性化的主要力量。所有建立在直覺基礎上的信念體系，都有一個共有的問題：當此類知識出現相互矛盾時，沒有一種機制可以判別孰對孰錯。因爲人人都憑直覺認爲自己是對的，但是當大家的感知有矛盾時，我們如何判斷誰是正確的？歷史告訴我們，這類矛盾通常，很不幸地，要靠權力爭鬥解決。

有些人錯誤地認爲，在心理學中使用操作性手段做研究是在剔除人性，因此我們應當改用研究者的直覺去了解人。心理學家布羅本（Donald Broadbent, 1973）指出：所謂研究的人性化，應該在於把關乎人的理論觀點，建立在大家都可觀測到的行爲，而不是理論家的直覺上：

> 除了觀察別人在特定環境中的所作所為，我們無從對別人做出判斷……實徵主義的方法提供一個擺平分歧的文明途徑。如果一個人拒絕這條道路，剩下來的唯一方法就是面紅耳赤的爭辯了。（p. 206）

因此，科學中推動人性化的力量是讓知識公開化，以便所有參與爭議的其他研究者可以用一種公認的方法平息爭端。記得在第一章我們提到的重驗性原則嗎？這個原則就是允許大家按照預先達成一致的平和操作機制，在衆多爭議的理論中進行篩選，找出最接近眞實世界的一個。而科學要能達到公開競爭性的關鍵，有賴於操作論這一原則。對概念所下的操作性定義，使得該概念公開化了；也就是說，任何人都可對其進行批判、驗證、改進或否定。

心理學概念不能以任何個別人士的定義爲基礎，因爲這種個人定義很可能是不尋常、古怪或模糊的。因此，心理學必須拒絕對任何概念作這類個人化的定義（如同物理學拒絕對「能量」作個人定義，氣象學拒絕對「雲」作個人定義那樣），並且必須堅持做可公開的操作性定義，以確保任何一個接受過適當培訓並擁有適當設施的人，都可以執行這些操作。拒絕個人定義，不但不會把大衆排除在心理學之外，還會把這個學科公諸於衆，如同所有學科那樣，讓大家都能使用、探索，並累積出可以爲所有人共用的知識。

只有當概念以操作性定義爲基礎，不理會本質論者對詞義的刨根問底，這種共用的知識才會出現，並用來解決人類的問題。舉例來說，蒙克（R. Monk, 1990）描述過一次大戰期間**創傷性休克**（wound shock）這一概念如何成爲醫學界的一道難題。當時，有些醫師確認，創傷性休克的表徵爲紅血球數量急增，並認爲原因在於血液中的血漿漏出，進入人體組織細胞。另一些人則認爲創傷性休克的表徵是血壓低、失血多、脈搏快。換句話說，醫師對這一概念的操作性定義很不一致（甚至帶

有個人定義的色彩）。一位在英國醫學研究學會（British Medical Research Council）工作的格蘭特（Grant）醫師於是建議：「應當把所謂『創傷性休克』這個概念丟掉。在不使用這個術語的情況下，對受傷的戰士進行詳細觀察和紀錄……由於缺乏大家都同意的診斷依據，以至於無法評估不同治療方法的功效。」（Monk, 1990, pp. 445-446）也就是說，這個概念沒有一個可以被視爲是通用的定義，所以使用它不但無益，還可能有害。

科學的概念會隨著研究的加深而產生操作性定義的改變。有時，這種改變會使概念在科學中的理解與行外人的理解產生矛盾。法博和丘奇蘭德（I. B. Farber & P. S. Churchland, 1995）圍繞著火這一概念論述了這種矛盾的情形。早期火的概念「不僅被用來指涉爲燃燒的含碳物質，還包括太陽及各種星體的活動（實際上是核子的熔解活動）、閃電（實際上是由電而產生的熾熱）、北極光（實際上是輻射光譜），以及螢火蟲的閃光（實際上是磷光）。在我們現代的概念系統中，因爲後面這些現象都不包含氧化過程，因此都不屬於木材燃燒的那一類。但是，另有一些現象原本認爲與燃燒沒有任何關係的，卻證實具有氧化過程，如生鏽、失去光澤以及新陳代謝等，所以可以與木材同屬燃燒類。這些現象在早期之所以不認爲與燃燒有任何相似之處，是因爲當時以能否感覺到熱度作爲區分是火或不是火的基本特徵。」（p. 1296）簡言之，氧化原則將生鏽與營火拉到一起——而將閃電與之分開——這對科學家來說，可能意味著界定火這一概念的新進展。但對外行人來說，反而讓他們感到迷惑，不知道火到底是個什麼東西了。

◎本質論式問題與對心理學的誤解

許多人想放棄操作論的另一個原因，是他們想求助於心理學去探究某些關於人類的根本問題，希望從中找到本質論式的答案。是由於心理學新近才從哲學中脫離出來？還是由於——相對於其他學科——大衆對心理學的了解實在是太少？究竟是哪個原因，目前還說不淸楚。就某種意義來講，其實是什麼原因並不重要，因爲最終結果都是一樣的：人們似乎期望心理學能夠明確地回答其他學科不能回答的問題。

回想一下本章開頭談及的那些問題：**地心引力**一詞的眞正含義是什麼？內在本質是什麼？在談到地心引力一詞時，我們到底指的是什麼意思？大多數人會意識到，這些問題關乎現象的本質，亦即其終極性質的問題，現在的物理學理論根本無法解決這類問題。任何一個熟讀近幾百年來關於物理學發展著作的人都知道，地心引力是一個相當複雜的理論概念。而且，其概念和操作之間的關係也不斷的變化。

然而如果把上述問題中的**地心引力**換成**智力**，突然間，怪事發生了。上述那些本質式問題一下子變得意義重大起來。這時候問起這些問題，不但變得理所當然，而且必須到有關人類終極關懷的本質領域中探尋答案。因此，每當心理學家像物理學家給一樣的答案——智力是個複雜的概念，它的意義由測量它的操作以及與其他概念之間的理論關係界定的——他就會被人看不起，並且會被指

責是在迴避作答。

這是心理學面臨的一個難題：大眾要求心理學回答那些本質論式的問題，而對其他學科卻通常沒有這樣要求。正是這些（心理學不能滿足的）要求，導致大眾對此領域所取得進步的輕視。幸虧大部分心理學家（就像其他科學家那樣）不去理會及迎合這些要求，只是專注於自己的工作，因而並未妨礙此一學科本身的發展。當然，這對大眾進一步了解心理學而言，確實增添不少阻力，所以當一個不了解情況的批評家宣稱心理學沒有任何進展時，大眾很容易就上當了。這些批評家的言論之所以極少遭到挑戰，恰恰反映了本書寫作的初衷：大眾對心理學的實際成就極度缺乏了解。仔細審視那些批評，不難發現通常歸結於一點：心理學至今沒有解決任何本質性的終極問題。對於這種指責，我們只能說抱歉，心理學確實沒能解決這類問題。但別忘了，其他學科也沒能做到。

沒有一個學科，包括心理學，能夠回答本質論式的問題。很多人可能對這一點感到不舒服。霍爾頓和羅勒（G. Holton & D. Roller, 1958）討論過，在得知物理學不能回答這類本質性問題時人們所反映的這種不安。他們談的是放射性衰變這一現象，是指一個放射性元素的原子數量的下降與時間呈指數函數關係。然而，該函數並不能解釋爲什麼會發生放射性衰變現象。要回答這個問題，可能必須用到另一個數學函數，但是那個數學函數仍不能回答大眾關於放射性衰變究竟是什麼的疑問。霍爾頓和羅勒告訴我們：「我們必須接受現代科學的局限性；它並沒有宣稱可以發現所謂『事情的最終眞相』。」（pp. 219-220）。正如科普作家賴特（Robert Wright, 1988）解釋的：

> 牛頓的地心引力理論也有些地方不盡人意……然而，究竟「在遙距動作」這一概念是否可能在現實生活中找到呢？……牛頓迴避了這樣的問題……從牛頓開始，物理學家都以此為榜樣……不再試圖解釋為什麼物體會遵循電磁學或地心引力法則。（p. 61）

同樣地，如果那些希望尋求關於人類本性的本質論問題的人，想求助於心理學，也必定要失望。心理學不是宗教，它是追求用科學的方法理解行爲各方面的一門寬廣的學科。因此，心理學目前對行爲所提出的解釋，還只是一些暫時比其他理論更契合的理論構想；將來這些理論構想肯定會被更接近眞理的概念替代。

◎操作論與心理學對問題的措辭

操作性定義的方法對評估心理學理論的可證僞性，是非常有用的工具。概念是否以直接或間接的可觀測操作爲基礎，是判別理論有無可證僞性的重要線索。那些沒有這種基礎的概念，通常是爲了配合蒐集後的資料刻意塑造出來的，爲的是避免理論遭到否定。因此，對待那些理論家不能給出直接或間接操作性關聯的不嚴謹概念，我們應抱持懷疑態度。

這與科學家所稱的**節約**（parsimony）原則有關。節約原則認爲，當兩套理論擁有同等解釋力時，更簡單的那套理論（含有較少概念和概念性相關）更可取。原因是，它含有較少概念性相關的理

論，在以後的測試中會更具有可證偽性。

對操作論原則的深入理解，有助於我們找到或辨別哪些是沒有科學意義的問題。例如，在我的檔案中有一篇來自合衆國際社（United Press International），名爲〈動物會思考嗎？〉的文章。它報導了最近對動物行爲所進行的一些實驗。文章所述的研究沒有任何不妥，但很顯然的，這個題目卻是在耍花腔。除非可以指明**思考**這一術語的操作性標準，否則題目中的問題就毫無科學意義。但是，在該文中並沒有提到操作性定義這個問題。許多最近的報紙刊文也存在類似的情況：「電腦能思考嗎？」同樣地，沒有對思考定出操作性標準，這類問題就無科學意義可言。不過，這些問題在鷄尾酒會上倒可以大派用場。

事實上，觀察人們就最後一個問題的爭論很有教育意義，因爲這種爭論使我們有機會具體見證，先前討論的對心理學所持的預設偏見問題。大多數人都有一個不想讓電腦也能思考的強烈預設偏見。爲什麼？主要原因在於：大衆對**思考**的想法，是與**人類**這一概念緊密糾結的，以至於許多人在情感上不能接受非人的東西也能思考（例如電腦，或看起來不像我們地球人的外星生命體）。

然而，儘管人們強烈排拒電腦能思考的想法，並不意味著大多數人仔細考慮過這個問題。對思考的定義，他們常常是糊裡糊塗地把絕大多數人（如嬰孩）都拉進去，而將所有電腦排除在外。沒有跟進目前人工智慧進展的人，對思考所提出的定義標準，有時聽來頗爲滑稽，因爲他們總是恰好選擇了那些電腦能勝任的作爲標準。例如，許多人提出思考是「從經驗中學習的能力」，那麼我們

只能告訴他們，用電腦運行的許多人工智慧系統都符合這一標準（Churchland, 1995; Clark, 2001; Elman, Bates, Johnson, Karmiloff-Smith, Parisi, & Plunkett, 1996; Pfeifer & Scherier, 1999）。然而，千萬不要小看預設偏見的力量，你可能以爲人們接下來會這樣反應：「噢，我不知道電腦能做這些事。既然電腦符合我提出的思考標準，那麼我不得不說至少有些電腦是能夠思考的。」錯了，通常大家不會做出這種理智而誠實的反應；更常見的反應是煞費苦心地重新另找些標準，希望電腦這次不能達到。

通常，思考的第二個定義標準是「創造性」（「想出人們認爲有用，但先前沒有其他人想到的東西——讓我們暫不考慮回答大多數**人**能否達到這一標準這個問題）。但即便大多數專家同意電腦符合這一標準（Boden, 1990; Pfeifer & Scherier, 1999），仍有人不肯承認機器可以思考的可能性。這時，他們往往不再求助另一個操作性定義，而是直接轉向爭辯：電腦不可能會思考，因爲「是人發明了電腦，並且爲它編寫程式；電腦不過是執行程式而已。」

儘管這種言辭早已成爲攻擊機器可以思考的陳腔濫調（McCorduck, 1979; Robinson, 1992; Woolley, 2000），它仍具有騙人的本事。預設偏見使許多人意識不到這些辯解與要討論的問題毫不相關。幾乎人人都同意，思考是自然界產生的一個過程。請注意，當我們在討論其他自然界的過程時，都沒有涉及「起源」的問題。試想一下食物加熱的過程，並這樣問：「是爐子在加熱嗎？」我們會不會說：「眞正在加熱的不是爐子，因爲它是**人**造出來的。因此只有說是**人**在加熱食物才有道理」？那麼輪到起重機時，我們又該怎麼講？「是起重機在舉重嗎？」我們的回答是不是還是「不是起重機

預設偏見：先入為主

在舉重，因爲它也是**人**造出來的，所以我們只能說是**人**在舉重」？當然不能這樣說。事物的起源與其執行什麼特定步驟的能力無關。思考的過程也是如此。一個實體能否思考，並不依賴於該實體是誰發明製造的。

針對人們無法理性地對待機器有可能會思考這一問題，電腦專家圖靈（Alan Turing）設計了一個測試電腦能否思考的著名實驗。這一實驗對於我們當前的討論尤爲重要，因爲圖靈設計的屬於**操作性**測試。圖靈以「我提議考慮這樣一個問題：『機器會思考嗎？』」爲開場白，開始他那篇著名的論文〈電腦和人工智慧〉（1950）。圖靈對電腦能否思考的探研，不是鷄尾酒會上的那種無聊閒扯，也不是本質論對**思考**的詞義作無休止的爭論，而是提出一個嚴格的操作性測試：如果電腦能夠進行智性對話，就可以說它有思考的能力。

圖靈這一提議的創造性在於，它既提出操作這個思考概念的方法，同時又防止了預設偏見的干擾。他嚴格地確認測試的每個環節，該測驗並不是簡單地把一部電腦終端機擺在屋裡，讓測試者用滑鼠和鍵盤與它對話，然後判斷電腦有沒有和這個人進行智性對話。圖靈的測試不是這樣做的，因爲他很擔心這一測試過程存在預設偏見的干擾。他確信，如果讓一個人坐在螢幕和鍵盤前，這個人顯然會意識到這些設備都是**機器**，那麼不管電腦是否真有思考的能力，他都會加以否認。因此圖靈建議，測試應控制那些與思考能力無關的外在因素。他的著名實驗設計是，讓受試者者同時與兩個鍵盤對話，其中一個鍵盤的一端是一部電腦，另一端是個活生生的人。當然受試者並不知道這兩

個鍵盤後面哪個是人、哪個是電腦。受試者的任務是在通過鍵盤與之對話後，判斷哪個鍵盤後面是電腦，哪個是人。如果受試者者不能以大於機率的正確率猜出哪個是人，我們就有理由推斷電腦具有同人一樣的對話能力——這正是思考的操作性定義。

圖靈的設計「與交響樂團選拔樂師的測聽面試之設置有異曲同工之妙。試聽面試時，在評審委員與面試樂師間要放一個不透明的屏風，評審要隔著這個屏風判斷樂師演奏的好壞。顯然，挑選演奏家，我們關心的應該只是他的音樂能力。性別、頭髮長短、膚色、體重等外在因素應該是毫不重要的，所以必須將它們摒除在外……圖靈認識到人們對智力的判斷，同樣會因對方是不是有柔軟的皮膚、一腔熱血，有面孔、手、眼，而產生偏見，這些特徵顯然都不是我們構想的智力之基本成分。」(Dennett, 1998, p. 5) 圖靈的測試告訴我們：要理性地討論心理學概念，而不只是反映我們自己對某個問題的偏見，操作性定義是必要的。

人們在討論人工智慧問題時使用的思維方式，恰當地說明了科學與非科學思考方法之間的差異。科學的方法是先發展出一個合理的操作性定義，再觀察我們可以得到哪些有關思考、電腦和人的結論。然而，預設偏見卻左右了一般大衆的思維方式。這些人對此一問題早就有先入爲主的結論，所以對機器與人的眞正差別到底在哪裡毫不感興趣。一旦定型，他們就會絞盡腦汁拚命找出各種理由以穩固這些想法，不讓它們改變。但是我們已經看到，正是這些預設偏見，加上非操作性的本質論心態，滋長了人們認爲無須任何必要的操作性標準，就能「知道」思考「眞正」是怎麼一回

事。正是這種心態，使得大多數人的直覺心理學理論無從證偽，因此也毫無價值。也正是這種心態清晰地向我們展示，爲什麼我們需要爲一門**科學**的心理學！

摘要

操作性定義是指，以可觀察、可測量的操作給一個概念下的定義。確保一個理論具有可證偽性的主要方法之一，就是弄清楚理論中的關鍵概念是否具有操作性定義，以及該操作性定義是否具有良好的可重驗性。操作性定義是科學知識得以公開並接受檢驗的一個主要機制。透過這樣定義的理論概念被公諸於衆，可以任由所有人重驗——不像直覺性定義是一種非實徵性的定義，只屬於個別個體所特有，無法公開地讓其他人驗證。

由於心理學使用的術語多源於日常話語，像**智力**、**焦慮**等，而許多人對這些術語的含義都有預先慣用的想法，所以對這些術語進行操作性定義的必要性，常常得不到人們的認可，認爲是多此一舉。其實，心理學與其他科學一樣，也是需要對術語進行操作性定義的。然而人們往往不這麼想，他們經常要求心理學回答一些本質論式的問題（關於概念的絕對、根本實質的問題），對其他學科卻不這麼要求。事實上，沒有任何一門科學能夠回答這類一成不變的終極性問題。心理學，與其他科學一樣，其操作性定義也是在不斷地修改中，以使其理論概念能更加精確地反映世界的原貌。

第四章　見證和個案研究證據

——安慰劑效應和神差蘭迪

讓我們在這裡將電視頻道切換到近十年來美國最受歡迎的一個脫口秀節目——《歐普拉脫口秀》（*Oprah Winfrey Show*）。今天的嘉賓是龐提菲科特（Alfred Pontificate）博士。他是伊底帕斯人類潛能研究所（Oedipus Institute of Human Potential）所長。主持人走到觀眾之間，要大家就博士剛剛提出的、目前尚存在不少爭議的出生順序理論進行提問，這個理論的基本思想是：一個人的生活歷程，早已在其與家人的互動中定型；而與家人互動的形式，又是由出生順序決定的。無一例外地，像這種討論的話題必然會從理論的探討，轉向對個別觀眾之重要個人事件的解釋上。對此，這位博士也欣然從命。

例如，有人問博士：「我哥哥是個自我毀滅式的工作狂，他不關心妻子，也不在乎家庭，只考慮與自己工作相關的事情。他患有潰瘍，還有酗酒問題，但他拒絕承認。他和他的家人已經兩年沒

像樣的一起度過假了。他逐漸在朝離婚的慘局走去，但是他好像根本不在乎他的婚姻。究竟爲什麼他會選擇這樣一條自我毀滅的道路呢？」

對這個問題，博士回應：「請問，他排行第幾？」

「哦，他是老大。」

「這就對了，」博士說：「這是很常見的。臨床中我總是見到這樣的個案。這類情況產生的原因是，父母把他們的希望和挫折都轉移給長子。這是一個無意識的過程，雖然父母沒有明說，但孩子還是會將這些希望和挫折內化成爲對自己的評價標準。我把這一無意識過程叫做期望上旋的動力過程，透過這一過程，父母的願望就會轉換成爲孩子對自己追求成就至病態的需要。」

儘管有時候，當嘉賓的觀點與觀衆的信念相左時，觀衆會提出一些尖銳的問題；不過如果所謂的行爲「專家」是一直在附和觀衆的傳統慣性思維，這種爭鋒相對的情況就不太會發生。不過，偶爾也還是會出現這類火爆的小插曲，這時場上的氣氛就熱鬧起來了。今天，現場就有這麼一位熱情、直率的提問者：「請等一下，博士，我哥哥也是長子，我父母把這個遊手好閒的傢伙送到哈佛讀書，而把我送去讀一個兩年制學校學牙醫。一年後，我父母心中的這位神童退學，去了科羅拉多的山區。我們最近一次見到他時，他正在那裡編籃子！你所說的長子現象我很不明白，好像與我哥哥的情況完全不同。」

觀衆席的氣氛頓時因爲這一挑戰而顯得緊張起來，但不打緊，這位博士總能給觀衆一個令人滿

意的解釋：「哦，是的，我見過很多像你哥哥這樣的個案。對，我在治療的過程中，經常會遇到這樣的人，他們的期望上旋動力過程已經中斷，因而產生一種無意識的願望，以阻止將父母的期望轉移到自己身上。這樣，個體的生活規畫就會朝著排斥傳統的成就標準的方向發展。」一陣肅然沈默之後，下一個「個案」又開始登場了。

以上所述，當然是大家相當熟悉的場景。這又是一個拉許問題的例子。這個出生順序「理論」是在沒有事例可以否定的框架下構想出來的。它是個不可證僞的理論，對它進行實徵研究是沒有意義的，因爲沒有任何事是不能用這個理論解釋，或排除於它的理論預測之外的。

然而，本章關注的問題並不是這個理論本身，而是用以支持這個理論成立的證據之性質及品質。龐提菲科特博士用他自己的「臨床經驗」或「個案研究」作爲支持其理論的證據。這在傳媒心理學是經常可見的。在脫口秀節目裡及書店一般大衆讀物的書架上，到處都是以創始人的臨床經驗爲基礎而成立的心理學理論。透過這些媒體管道，一些展現在大衆面前的治療法都僅僅是出自接受過該項治療，並認爲這種治療有效的個體的親身口述。這種口述通常稱爲「見證」（testimonials）。在這一章，我們主要想說明一個道理，提醒想知道如何善用心理學知識的人：用個案研究結果或見證敘述作爲評價心理學理論和治療效果的依憑，幾乎可以說是完全沒有價值的。

以下，我們將闡明爲什麼這些證據不適合用來支持理論及療效，並討論個案研究在心理學研究中的作用何在。

個案研究的作用

個案研究資訊的價值，很大程度上取決於某一領域科學研究的進展。在對一些問題之研究的早期階段，從個案研究或臨床經驗中獲得靈感與啓發是非常有用的，它們提示我們哪些變數值得進一步研究。個案研究對心理學開闢新的研究領域起過非常關鍵的作用。皮亞傑（Jean Piaget）的工作就是人盡皆知的一個例子。他透過研究，提出兒童的思維不是成人思維的簡易版或初級版，而本身是自成一體的想法。皮亞傑關於兒童思維的這一推測，有些已經得到證實，但是很多到目前還沒有證實（Bjorklund,1995; Goswami, 1998）。不過，這裡討論的重點不是他有多少設想得到證實，而是要理解：皮亞傑的個案研究沒有**證實**任何東西，它只是爲發展心理學提供非常多有待進一步研究的亮點。眞正爲皮亞傑從個案研究中提出的假設提供證實或證僞之證據的，是隨後進行的一些相關研究和實驗研究（我們將在第五和第六章談到）。

然而，當我們從科學研究的早期階段過渡到比較成熟的理論檢驗階段，情況就徹底改變了。個案研究在科學研究的後期是發揮不了什麼作用的，因爲這時主要的研究工作是檢驗特定的理論，個案研究不能提供證實或證僞這些理論的證據。原因是：個案研究和見證敘述都是所謂的孤立事件，缺乏必要的比較資訊，以有助於排除對同一事件或例證的其他可能的合理解釋。

見證敘述之所以——和前面的個案研究例子一樣——是孤立事件，因此不能驗證對錯，是因爲

見證叙述通常只包括支持某一理論、治療或事件的話語。如此，問題來了，如果每個試驗療法都可以累積一些支持其療效的話語，那麼，我們又如何可能用之判定某個療法是好是壞。正如心理學家尼可森（Ray Nickerson, 1998）在評論我們常常用以自欺的認知過程時所說的：「江湖郎中的騙術往往得逞，是因爲他們隨時可以找到一些病患願意爲他們做見證，這些病患總是滿懷誠心地告訴他人，自己的確在治療中有所改善。」（p. 192）例如，一些賣潛意識自助錄音帶（那些用一種低於聽覺閾限的訊號製作的錄音帶），往往以大量的用戶給出見證證據，聲稱可以提高人的記憶力、甚至自尊。然而，在嚴格控制條件下的研究顯示：這類錄音帶對記憶力或自尊沒有任何改進（Greenwald, Spangenberg, Pratkanis, & Eskenazi, 1991; Moore, 1995）。

理解理論檢驗最關鍵的，是要認識「其他可能的合理解釋」這一概念：一個事件或現象通常可以用許多可相互替代的理論解釋。實驗設計的目標，就是組織安排這一事件或現象的出現，使其只能用某特定理論解釋，其他理論則解釋不通。正如我們在第二章所說的，只有利用精心安排的實驗程序，排除其他合理解釋的可能性，我們所得的實徵資料才是有價值的，科學研究才會有進步。科學可以說是爲各種相互競爭的理論或觀點提供一個自然選擇及淘汰的途徑。有些解釋經過實徵檢驗存活下來，另一些則遭淘汰。保留下來的可以說是比較接近眞理的。但這是個慢工出細活的工作，各種理論觀點必須反覆接受審查，以便發現哪個最接近眞理。同時，這必須是一個「你死我活」的競爭過程：爲支持某一特定理論所蒐集的證據，不能同時可用以支持其他同等可能的合理解釋。基

於這一原因，科學家在他們的實驗中設有控制組，或稱爲比較組，以求得比較資訊。這樣做的目的，是在對控制組的結果與實驗組進行比較時，其他可能的合理解釋可以排除。實驗設計如何做到這一點，是後面幾章我們要談的內容。

個案研究和見證叙述因爲沒有給出比較資訊，只代表一些偶發的孤立現象。它們缺乏必要的比較資訊，以證明某一特定理論或治療更具優勢。因此，引用見證叙述或個案研究支持該**特定**理論或治療，是個錯誤的舉措。這樣做的人，如果不向大衆挑明他們提供的這類所謂證據，很可能許多其他的合理解釋也解釋得通，就是在誤導群衆。總之，對現象的這種單一孤立的論證，具有極高的誤導性。我們下面將用安慰劑效應（placebo effects）更具體地說明。

見證叙述爲什麽没價值：安慰劑效應

實際上，醫學和心理學實施的每種治療都能獲得一些支持者，總有人誠心誠意地證實其效果。在醫藥科學的記載中，有人見證過，豬牙、鰐魚糞便、埃及木乃伊製成的粉，以及其他更奇怪、更富想像力的「藥物」都具有療效（Harrington, 1997; Shopiro, 1960）。但實際上，我們早就知道：僅靠暗示某一治療正在進行中，就足以讓許多病患感到病情好轉。

不管治療是不是眞的有效，人們通常都傾向於報告它們有效，這就是著名的**安慰劑效應**（Chris-

tensen, 2001; Dodes, 1997; Ernst & Abbot 1999; Harrington, 1997; Kirsch, 1999; Russo, 2002）。在電影《綠野仙蹤》（*The Wizard of Oz*）中，這個概念得到很好的展示。巫師**實際上**沒有給錫人一顆心、給稻草人一個腦、給獅子以勇氣，但是他們都感到比以前更有愛心、更聰明、更具勇氣。實際上，由於直到近一百多年，醫藥科學才發展出較多具有確鑿療效證據的治療，有人曾經這樣說：「在這個世紀以前，醫學史不過是安慰劑效應的歷史。」（Postman, 1988, p. 96）

我們可以透過考察生物醫學研究來說明安慰劑這一概念。在生物醫學研究中，所有的新藥研究程序必須包括對安慰劑效應的控制。一般來說，如果在一組病患身上試驗一種新藥，就要組建一個患同樣病症的對照組，給他們服用不含該藥的等量藥劑（安慰劑）。兩組病患都不知道他們吃的是什麼藥。這樣進行比較時，安慰劑效應——即給予病患任何一種新的治療都會使他們感覺好些的這種傾向——就可以得到控制。僅僅報告百分之幾的病患吃了新藥後症狀得以緩解是不夠的，因爲如果我們沒有控制組的緩解百分比資料，就不知道報告症狀緩解這百分之幾的病患，是安慰劑效應還是藥物本身的效果。對各種疾病患者的研究（參見 Bok, 1974）顯示，大約有35%的人吃了安慰劑後，病情也獲得令人滿意的緩解（同時參見 Roberts et al., 1993; Russo, 2002）。最近的一個研究認爲：與常用抗憂鬱藥百憂解（Prozac）相對應的安慰劑效應，是該藥本身效果的兩倍（Bower, 1996b）。甚至有人發現，安慰劑效應可以強到令病患上癮的地步（Bok, 1974; Ernst & Abbot, 1999），從而必須依靠吃安慰劑維持健康！

當然，在藥物療效的對比研究中，對安慰效應的控制，事實上並非給病患一個完全沒有療效的藥片。相反地，它是當前認爲能控制該症的最好藥片。實驗比較所要揭示的問題是，新藥是不是比當前最有療效的藥更好。

安慰劑效應在心理治療中也很常見（Wampold et al., 1997）。許多有輕度和中度心理問題的人，在接受心理治療後，說他們的情況好轉。然而控制研究證明：此一康復比例中有相當一部分人，是由於安慰劑效應和單純的時間推移這兩個因素共同作用的結果，而這裡的時間推移，通常稱爲**自然康復現象**。大多數有療效的治療，都是治療效果和安慰劑效應的結合。正如多德茲（J. E. Dodes, 1997）所說：「即便非常嚴重的疾病，也有惡化與康復交替出現的過程；關節炎和多發性硬化是最常見的例子。甚至連有些人患有腫瘤，也會莫名其妙地消失。」（p. 45）不過他也提醒大家：有安慰劑效應，並不意味著病患的問題是無中生有的。和一般看法認爲安慰劑是無害的正相反，他提出警告：「安慰劑效應往往會證實和（或）強化人們懷疑自己有某一疾病的想法，從而『誘發』出慢性疾病，使他們變成毫無科學根據、只靠安慰劑效應從事治療的江湖郎中之常客。」（Dodes, 1997, p. 45）

在心理治療療效的研究中，通常很難決定該如何對待服用安慰劑的控制組，才能讓他們給我們所需的比較資訊。不過，讓我們暫且不談這個複雜性問題，而先理解一個更基本、更重要的問題：研究人員爲什麼要區分治療效果與安慰劑效應和自然康復現象。保羅（Gordon Paul, 1966, 1967）的一項對心理治療療效的研究，提供很好的例子，讓我們看到這類研究能夠告訴我們什麼。保羅研究了

幾組在公共場合說話會感到不安及焦慮的學生。其中實驗組接受專門針對說話焦慮問題的減敏感治療，有85%得到明顯改善。安慰劑組吃的藥被告知是一種強力鎮靜劑，但實際上只是小蘇打膠囊；這一組中，50%的人說有明顯改善。第三組沒有接受任何治療，也有22%說明顯改善。從這一研究中我們可以看到，這種心理問題的自然康復率爲22%，由於安慰劑效應而得到改善的百分比爲28%（50%減22%），而減敏感治療的療效則爲在安慰劑效應之上的35%（85%減50%）。

與保羅的研究結果一樣，其他研究也表明：心理治療的療效確實優於只用安慰劑所可能產生的效果（Lipsey & Wilson, 1993; Nathan & Gorman, 1998; Prioleau, Murdock, & Brody, 1983; Smith, Glass, & Miller, 1980）。但是使用安慰劑爲控制組的實驗也表明：如果我們僅僅用實驗組與控制組病患自我報告感覺有改善的人數之百分比的差異，會嚴重高估該治療的實際療效。這裡的問題在於，見證敘述實在太容易得到了。康乃爾大學（Cornell University）心理學家基勒維奇（Thomas Gilovich, 1991）曾指出：「由於人體有很好的自癒力，即使醫師什麼管用的事都沒做，病患中有很多也會感到病情好轉。因此，當因自癒令治療率的基線變得很高時，甚至一個完全沒用的治療也會顯得很有效。」（p. 128）總之，只要我們進行醫療干預，安慰劑效應就無時無刻不在，**不管這種干預本身是否眞的有效**。由於安慰劑效應如此強烈，哪怕治療方法再荒唐滑稽，只要它是向一大群人實施，總有幾個人願意出來見證它的療效（清早擊頭治療法——每天早起敲打你的頭部，你會一天精神百倍！郵寄十塊九毛五美元給我，你就可以得到這支特製的、經醫學測試的橡膠槌）。

對不起，這麼嚴肅的事，我們眞的不應該開玩笑。毫無理由地依賴見證敘述和個案研究證據，很可能帶來災難性後果。例如，爲妥瑞氏症做出科學概念化——將之界定爲器官障礙——的研究小組（Shapiro et al., 1978，見第二章），其成員曾經指出，由於人們不恰當地依賴個案研究證據，使得對該病的病因一直停留在不可證僞的心理分析理論框架內打轉，阻礙了對病理進行眞正科學研究的進展。英國科學雜誌《自然》駐華盛頓記者巴帝安斯基（Stephen Budiansky, 1984）在總結醫學領域中的這種現象時，曾說了下面這段話，更強化了我們前面幾章討論過的關於什麼是科學的許多要點：

> 科學迴避所有個人化的東西。雖然這一偏向常歸因於科學家應有的冷靜，但事實上，它確實是二十世紀人類智慧的偉大戰利品之一。科學家深知，要尋找自然世界的原因和結果，就必須排除個人的感覺及經驗。健康科學曾為此吃盡苦頭；人們生病有許多原因，大多數情況下，人們無論在接受什麼「治療」後，都會有所好轉。所以，看來永無止盡的各類庸醫庸術，憑藉著那些對療效滿意患者的見證敘述，總是廣受人們愛戴。這些就是活生生的例子，說明要我們跳出個人經驗的束縛有多麼困難。（p. 7）

《新英格蘭醫學期刊》（*New England Journal of Medicine*）的一篇社論，反映了醫學界對個案研究及見證敘述在醫學研究中可以發揮的作用：「比方說，如果我們雜誌接受一篇論文，是敘述一個胰腺癌病患在服用大黃（rhubarb）後康復……我們可能會作爲一個案報告發表，但是我們之所以發表

並不是宣告它爲一種新療法，而僅僅是建議它是一個值得用正規的臨床實驗測試進行驗證的假設。與此正相反，那些講述各類祕方之療效的文章（通常發表在通俗的書籍和雜誌上），則沒有此種聲明，因此文章本身也就不足以作爲對那些療效的支持文獻了。」（Angell & Kassirer, 1998, pp. 839-840）

「鮮活性」問題

前面我們已經闡明，因爲有安慰劑效應，致使見證敘述無法成爲有用的科學證據。能指出這一點固然好，但還不夠，要人們摒棄見證式證據還須掃除另一個障礙。社會和認知心理學家都曾對人類記憶和決策過程中出現的鮮活性效應（vividness effect）做了許多研究（參見 Baron, 1998; Kunda, 1999; Ruscio, 2000）。當面臨解決問題或作決策的情境時，人們會從記憶中提取與當前情境有關的資訊。因此，對解決當前問題而言，愈是容易獲取的事實，愈會被人們用來解決問題或作出決策。而資訊的鮮活性，就是強烈影響資訊之可獲取性的一個因素。

問題來了，因爲沒什麼比人們眞實地敘述他們看到的東西或經歷過的事，更具有鮮活性、更能打動人心。個人見證的鮮活性經常使得其他更可靠的資訊黯然失色。有多少次，我們在購物前雖然仔細蒐集了大量不同產品的資訊，卻在最後時刻，僅因一個朋友或一個廣告而改變自己最初的購買決定？購買汽車是個典型的例子。大概在翻看了《消費者報導》（*Consumer Reports*）中的數千份消費

者調查之後，我們終於決定購買一輛X車。又參考了幾本汽車雜誌，裡面的專家也推薦X車，更堅定了我們的選擇——直到一次聚會中我們遇到一個朋友，他說他朋友的朋友買了一輛X車，結果是一輛爛車，老有毛病，花了幾百美元維修，而且決定再也不買這個牌子的車了。單單這樣一個事例本不應該那麼嚴重地影響我們對X車的觀點，因爲我們是在蒐集了針對數千名用戶所做的調查報告和衆位專家的評判之後，才決定要買X車的。然而，我們之中究竟有多少人會不受這一見證敘述的影響呢？

買車的事例說明了：鮮活見證證據所產生的問題不只存在於心理學領域中。鮮活性影響人們觀點的事例處處可見。記者強森（Haynes Johnson, 1991）曾記載美國雷根總統（Ronald Reagan）如何認識到愛滋病之嚴重性的事例：

> （雷根）直到一九八五年才認識到愛滋病的嚴重性，當時他看到一則新聞報導，演員赫德森（Rock Hudson）死於這一病症。此時愛滋病已出現五年之久，數以千計的美國人遭到感染，關於它的話題也早已傳遍街頭巷尾。但雷根是看到赫德森的死亡報導時，才要赫頓將軍（General Hutton，以前做過醫師）給他講講這種疾病。（p. 454）（譯注：赫德森是曾任好萊塢演員的雷根之好友。）

簡言之，新聞報導和統計資料都顯示愛滋病死亡人數不斷增加，但是這些資料並沒有能夠吸引

總統的注意；而他所認識的朋友中，僅僅一個人死於此症就足以引起他的注意。同樣，作家路易斯（Michael Lewis, 1997）也講述了著名保守派時事評論員威爾（George Will）——一個堅決反對強勢政府干預老百姓生活的人——是在他家門外看到撞車造成死亡事件後，才發表專欄文章呼籲政府要強制所有汽車安裝安全氣囊。

試想一下，一個週五早晨你在報紙上看到一個大標題：「巨型客機墜毀：四一三人喪生」。你可能會想：天啊，多可怕的事呀！怎麼會發生這樣悲慘的事情！再試想，過幾天之後到週四，你一起床看到報紙上說：「另一架巨型飛機遇空難：四四二人死亡」。那時你可能會想：唉呀！不可能吧！多可怕呀！我們的空中交通系統到底出了什麼問題？接著再試想，又到下一個週五，報紙上又出現：「第三個撞機悲劇：四三一人命喪黃泉」。這時，恐怕不僅是你，全國都會震動起來，聯邦政府會被要求儘快調查此事，飛機會被停飛，許多調查委員會將設立，並引來一大堆的訴訟案件。《新聞週刊》（*Newsweek*）和《時代》雜誌（*Time*）會以此為封面報導，連續幾天這件事都會是電視新聞節目的頭條，電視紀錄片也會對此作詳盡的分析。一時間，硝煙萬丈。

請注意，這裡敘述的現象並不是虛構的，它所帶來的問題是真的。每週**的確**有一架巨型飛機墜毀。好吧，也許不是一架巨型飛機，而是許多架小飛機行不行？不行，好，不是小型飛機，而是小型交通工具可不可以？這種小型交通工具叫汽機車。全美國每週大約有八四三人死於交通事故（根據美國國家安全委員會〔National Safety Council〕二〇〇一年的報告，每年共四萬三千五百人）。死者足以坐滿

一架巨無霸噴射機，這總可以了吧！

每週死在高速公路上的人，與一架巨無霸噴射機墜毀的死亡人數相當，**卻沒有引起我們的注意！**這是因爲媒體沒有以「正如一架滿載乘客的巨無霸噴射機墜毀」這類鮮活的方式呈現給我們。所以，**每週**死於小汽車交通事故中的四五七人（加上**每週**死於卡車和摩托車事故的二六三人），對我們來說，沒有鮮活性，我們不會在餐桌上像談論飛機墜毀、衆人死亡那樣地談論他們。不像如果飛機眞的每週墜毀、每次死亡四百人，我們一定會討論空中交通系統的安全性那樣，我們也沒有爭辯汽車行駛的安全性和必要性。媒體沒有鮮活地給我們呈現這四五七人的死亡，因爲這些交通事故不是發生在同一時間、同一地點，因此媒體只能（偶爾地）給我們提供一些數字（如每週四五七人）而已。這一現象**應該**足以引發我們思考，但卻沒有。根據調查資料，駕車是一項極端危險的活動（Lichtenstein, Slovic, Fischhoff, Layman, & Combs, 1978; National Safety Council, 1990, 2001）。然而，至今從未見有就這一活動所帶來的危險性——相對於帶來的便利——進行任何全國性大辯論。難道這裡說的死亡數字，就是人們爲了享有當今的郊區生活，因而需要大量駕車往返於工作及居住地之間，所應付出的昂貴代價嗎？我們從來沒有提過這類的問題，因爲我們從來沒有認識到這是個問題。因爲它付出的代價沒有像空難那樣，以鮮活的方式呈現給大家。

請再試想下面的例子，看它有多滑稽。你的朋友開了二十哩的車送你去機場，你從那裡乘飛機到一個七百五十哩以外的地方。離別時，朋友可能對你說：「一路平安。」這句話傷感又頗具諷刺

意味。因爲**他在回家的二十哩車程中死於交通事故的機率，比你飛的七百五十哩航程高出三倍有餘**（National Safety Council, 1990）。「A祝B旅途平安」事實上是一個非常不理智的祝福，因爲A的處境事實上比B更危險。然而，受到鮮活性效應的影響，讓我們覺得A給B的祝福是合理的（Sivak & Flannagan, 2003）。

科爾（K. C. Cole, 1998）在一次環球航空公司（TWA）空難後所做的民意調查中發現，如果飛行安全係數能夠提高，絕大多數受調查者願意爲一張雙程票多付五十美元。然而，同樣的這些受訪者卻拒絕支付比五十元少得多的錢，去購買一些可以帶給駕駛者更多保障的車內安全措施。同樣地，很少人知道：他們在駕車十哩去買一張樂透彩券時，死於交通事故的機率比中獎的機率高出十六倍！（Orkin, 2000）

依據大衆媒體展示的鮮活形象，誤導人的個人判斷可謂隨處可見。一項針對父母的研究調查報告（McDonald, 1990；同時參見 Cole, 1998），要他們舉出最擔心孩子遭遇什麼危險。結果顯示父母最擔心的是孩子遭綁架，這事件的機率是七十萬分之一。與之相對照，孩子在車禍中身亡的危險父母則不太擔心，然而它比遭綁架的可能性高出**約一百倍**（Paulos, 1988）。顯然，遭綁架的擔憂大部分是媒體渲染的結果。這項研究結果說明，由於鮮活性效應帶來感知上的扭曲，「美國父母擔心的竟是一些不大可能發生的事」（McDonald, 1990）。一個研究人員哀歎道：將擔心聚焦在「當下流行」的事件上，必然會誤導父母的注意力，致使他們「忽視了更有可能造成壞影響的事件，例如學業成績、看

電視的壞習慣、吸毒和駕車安全等。」（McDonald, 1990）

作家博伊爾（Peter Boyer, 1999）曾描述，維護攜槍自由的說客如何以同樣的方式助長美國人對危險的錯誤認識。這些人試圖讓大衆的注意力集中在「非法入侵者」破門而入的鮮活案例上。他們那並不難懂的潛台詞是：這時，如果有一把槍保護自己，你就沒危險了。但是博伊爾（1999）指出，極具諷刺意味的是，槍支產業試圖把問題的焦點放在「壞人手中的槍」上，而實際的統計數字卻顯示：在現實生活中，問題往往出現在「好人手中的槍」上。這是因爲犯罪並不是這個國家大多數槍擊死亡案的主要原因，事實上死於用槍自殺的人比槍殺案多得多。多數槍擊死亡可歸因於槍支走火和自殺。這正是爲什麼研究結論是：把槍帶回家反而增加了家庭的危險（Miller, Azrael, & Hemenway, 2002）。

◎單一個案的壓倒性影響

雖然人們有許多正確資訊可以幫助自己作判斷，但是他們往往不顧這些資訊，而被一個單一的、突出的例子左右。對人的這種傾向，心理學家曾展開深入的研究。威爾遜和布瑞奇（T. D. Wilson & N. Brekke, 1994）證實了鮮活性的危害，以及它對消費者行爲的實際影響。他們調查兩個不同品牌保險套（A牌和B牌）的兩類資訊對人們產生的不同影響。一類資訊是《消費者報導》雜誌上的一次調查及分析，另一類是由兩名大學生敘述他們對某一品牌的偏愛。首先，威爾遜和布瑞奇

調查一組受試者：詢問他們較願意接受哪一類資訊的影響。超過85%的受試者說，他們會選擇《消費者報導》，而不是那兩個大學生。然後，威爾遜和布瑞奇選擇另一組與前一組相匹配的受試者進行實驗，他們告訴受試者可以免費得到自己選擇的保險套，而且受試者在選擇的過程中，可以參考其中一類的資訊，或兩類都考慮——亦即綜合《消費者報導》雜誌的調查分析和兩個大學生偏愛的敘述。儘管先前只有不到15%類似的受試者願意接受兩個學生之偏愛的影響，但是這一次，77%的受試者選擇兩類資訊都要。很明顯地，即使受試者不認爲他們應當受見證叙述的影響，但最終還是難以抵擋見證資訊的誘惑。並且，他們也的確受到見證資訊的影響。當受試者選擇同時看兩類資訊，而兩類資訊推薦的保險套又不同時，31%的受試者選擇大學生見證叙述中推薦的，而不是《消費者報導》的。

另一個人們對鮮活軼聞做出特別反應的例證，是關於一九六〇年代中後期媒體對越戰的報導。隨著戰爭沒完沒了，士兵死亡名單彷彿永無盡頭地增長，媒體不得不採取每週彙報上週美軍的死亡人數。週復一週，這個數字大致總在二百到三百人之間，漸漸大衆似乎也習以爲常了。然而，有一週，一份重要雜誌用連續幾頁的篇幅，將上一週戰死人員的照片全部刊登出來。這時大衆非常具體地看到，大約二百五十個活生生的人在過去一週內死去。結果，對這場戰爭造成的損失之強烈抗議爆發了。二百五十張照片所產生的影響是每週數字遠遠不及的。但是作爲一個社會的成員，我們應該克服這種不相信數字而必須親眼目睹才相信的習慣。絕大多數影響我們社會的複雜因素，都是只

有靠數字才能捕捉的。只有當大眾學會重視以數字形式表達的抽象材料，如同重視形象表達的材料時，自己的觀點才不會隨螢幕上閃過的最新形象而左右搖動。

不僅大眾跳不出鮮活性這一因素的困擾，經驗豐富的臨床心理學家和醫師在做判斷時，也總要儘量抵制單一個案的壓倒性影響。作家拉索（Francine Russo, 1999）曾描述維吉尼亞大學（University of Virginia）腫瘤學家安德森（Willie Anderson）陷入的兩難困境。安德森一直提倡用控制實驗法，並且時時招募及儲備一些病患做有控制的臨床測試。但是他知道在面對有些突出的個別病例時，情感因素會影響他的判斷。這時，他不得不與自己的這種對單一個案的情感反應作戰。儘管力求從科學角度出發，他還是承認「當一個真實的人眼巴巴地望著你的眼睛時，你會被他們的希望以及自己對他們希望的希望包圍，這的確**很難過**。」（p. 36）但安德森知道，有時對病患最好的事，就是不去看「望著你眼睛的那個真實的人」，而是根據最佳證據做。所謂最佳證據，是來自有控制的臨床實驗（我們將在第六章表述），而不是對那望著你眼睛的那個人所做出的情感反應。

評價證據時，鮮活性問題是我們要面對的一個難題。在到處都滲透著媒體形象的環境中，要整個社會不受形象的影響，只依據可靠的證據解決社會問題，是愈來愈難了。作家格拉斯納（Barry Glassner, 1999）講述過一個大家非常熟悉的例子。一九九五年在《歐普拉脫口秀》的一次節目中，討論圍繞著某一外科手術展開（爲了不替鮮活效應助紂爲虐，這裡不點出這種外科手術是什麼），這種手術曾經引起一些爭論，因爲有人聲稱曾被它所傷害。但是來自梅約醫學中心（Mayo Clinic）、哈佛大學

（Harvard University）和密西根大學（University of Michigan）的研究證據都顯示：這種手術程序一般來說沒有危險性——然而正在此時，一名聲稱受過這個手術傷害的婦女突然從觀眾席中跳出來，大叫道「我們就是證據。你們應該研究的是我們這些坐在這裡的受害者。」（Glassner, 1999, p. 164）你認爲數百萬電視觀眾會對什麼東西記憶猶新？——梅約醫學中心的研究？還是大叫著自己受過傷害的婦女？

總之，由於依賴見證證據而產生的問題經常存在，這種證據的鮮活性經常掩蓋更可靠的資訊，從而混淆視聽。但是心理學教師擔心，僅僅指出依賴見證證據的邏輯錯誤，並不足以令學生深刻理解這類資料的陷阱。那麼，我們還能做些什麼？還有什麼其他方式可以讓大眾了解這一概念嗎？所幸倒眞還有一種方法——一種略略不同於學術研究的方法——可以達到我們的目的。這一方法的重點就是用鮮活性對抗鮮活性。讓見證敘述搬石頭砸自己的腳，讓見證方法自身的荒謬來毀滅它自己。這一方法的施用者之一，就是我們獨一無二、毋庸置疑的神差蘭迪（Amazing Randi）！

◎神差蘭迪：以其人之道還治其人

蘭迪（James Randi）是一名魔術師和什麼都會的人，曾經獲得麥克阿瑟基金會（MacArthur Foundation）頒發的「天才」獎。多年來，他一直致力於教育大眾一些進行批判思考所需的基本功夫。神差蘭迪（他的藝名）以拆穿許多自稱通靈人的騙術來教育群衆。他揭發過無數假稱自己有通

靈能力的魔術師及大法師，最有名的是一九七〇年代揭露通靈超級巨星格勒（Uri Geller）詐術的那一次。格勒以即席表演的方式大事鼓吹他的通靈力量，將全世界的媒體焦點全部吸引過來。幾大洲的報紙、電視和主要新聞雜誌，無不針對他的奇術爭相報導。蘭迪則察覺並披露出，格勒用以表演通靈「絕技」的祕訣，竟然是一些非常普通及簡單得讓人不好意思說的魔術小把戲，包括彎曲鑰匙或湯匙、啓動鐘錶等等。對一個優秀的魔術師來說，這是再平凡不過的手藝了。在格勒的騙術不停曝光之後，蘭迪繼續不斷地用他那非凡的才智，揭露像超感官知覺、生物節律、通靈外科手術、外星人、空中漂浮及其他各式各樣的僞科學，從而維護了大眾知的權利（Randi, 1986, 1987, 1995; Sagan, 1996）。

蘭迪許多較次要的興趣中，包括向大眾展示任何荒謬的事情或無中生有的說法，都非常容易蒐集到見證證據（Sagan, 1996）。他的方法是讓人們落入自己用見證所設的陷阱中。蘭迪充分利用美妙的美國文化窗口——脫口秀，經常隱藏自己的眞實姓名，假扮他人出現在這類節目中做特別來賓。在幾年前紐約的一次節目中，他告訴觀眾，他那天早上駕車經過新澤西州時，看到一個橙色的V形物體從北面飛過頭頂。幾秒鐘內，蘭迪說：「電視台的電話接收機燈就像電子耶誕樹一樣的全亮了。」節目播出後，不斷有目擊者打電話進來證實他們也看到這一奇異景象，並補充許多蘭迪在節目中「忽略」的細節，其中包括其實不只一艘「飛碟」經過，而是很多艘。他們當然都沒想到，原來該「奇象」只是蘭迪編出來的。這個小把戲，一下子就說明了個人「發生了某一件事」的經歷敘

述是完全靠不住的。

在加拿大溫尼伯（Winnipeg），蘭迪以「天文星相學家」的身分出現在一個電台節目上。一週前節目主持人曾邀請聽衆將他們的筆跡和出生日期提供給蘭迪，好讓他從中選出三位，並與他們保持聯繫，以便從「生辰八字」看看他們是麼樣的性格。他大爲成功。在一個1到10分準確度量表上，這三位聽衆對他的預測的準確度評分是9、10和10。蘭迪最終還是向聽衆抖出他的預測祕訣。原來，他只不過是一字不露地按照星相學家奧瑪（Sydney Omarr）不久之前在一個電視節目中爲三位觀衆「預測」的原文，重讀一遍而已。

在另一次電台節目中，蘭迪演示了另一種非常盛行的僞科學：生物節律（Hines, 1998）。一位聽衆同意每天記日記，以供將來與一個爲她準備的、爲期兩個月的生物節律圖進行對照。兩個月後，那位婦女打電話回來告訴聽衆：應該認眞對待生物節律，因爲她的圖表準確地預測了她的實際行止，正確率超過90%。蘭迪不得不告訴她，他的祕書犯了個愚蠢的錯誤，把另一個人的圖表給了她。之後，那位婦女同意看一看自己眞實的圖表究竟怎樣。於是一份圖表很快又寄給她，並要她再打電話來。幾天後，那位婦女回電話說她鬆了一口氣。原來，眞正屬於她的圖表也相當準確，甚至比原來那張錯的更準確！殊不知，在接下來的一次節目中，蘭迪發現又搞錯了，第二次送給那位婦女的竟然是蘭迪祕書的圖表，不是她的！

蘭迪的生物節律和天文星相學家詭計，實際上都是稱爲巴楠姆效應（P. T. Barnum effect）的一類

現象的例子。（巴楠姆是著名的狂歡節目和馬戲團團主，他曾發明這樣一句話：「容易上當受騙的人每秒出生一個。」）對此一效應，心理學家曾進行深入的研究（如 Dickson & Kelly, 1985）。他們發現，絕大多數人認爲泛化的人格特徵描述都是正確的，並把這些特徵視爲自己獨有的。巴楠姆效應讓製造見證式證據變得非常容易，這當然正說明了爲什麼這類證據是毫無價值的。心理學家的研究發現，有些詞句是絕大多數人認爲適用於自己身上的，任何人都可以把這些詞句貼在他們的「客戶」身上，算是對他們個別的心理「分析」，而客戶都會對這些個別化的「人格分析」之準確性印象深刻，完全不知道其實給每個人的詞句都是一樣的。當然，巴楠姆效應是在相信手相和占星術的人身上才會奏效（Kelly, 1997, 1998）。

蘭迪這些小詭計的目的，是想給大衆上一課，教育他們見證證據是沒有價值的。他毫不例外地演示出，不管一個人說的是眞是假，如果他想製造一些見證證據證明自己說的是眞的，那是再容易不過了（Randi, 1983）。由於這個原因，透過見證支持自己提出的特定理論或說法是毫無意義的；只有在控制條件下所做的觀察中獲取的證據（將在第六章介紹），才足以檢驗這些理論及說法。

見證爲僞科學大開方便之門

有時有人會說，類似剛才討論的種種僞科學，不過是人們點綴生活的一種方式，大家笑一笑，

無傷大雅。老實說，誰會對這些事認眞呢？不就是有幾個人在異想天開，而另外幾個人從中賺幾個臭錢嗎？

然而，對此一問題進行全面考察就不難發現：僞科學盛行對社會的危害比人們想像的大得多。第一，相信僞科學會導致個人悲劇。看看薩林斯（Rhea Sullins）的個案。薩林斯是美國自然保健協會（Americna Natural Hygiene Society）主席的七歲女兒。這個協會主張人們可以吃水果和蔬菜取代用藥和傳統治療。薩林斯生病後，「她父親讓她禁食十八天，其間只給她喝水。接下來的十七天則只准她喝少量果汁。」（Gilovich, 1991, P. 6）可悲的是，薩林斯最終死於營養不良。更不幸的是，像薩林斯這樣的例子比比皆是（參見 Kaminer, 1999; Shermer, 1997; Stanovich, 2004; Sternberg, 2002）。

在一個複雜、高科技的社會裡，僞科學的影響可能由於某些涉及千千萬萬人利益的各種決策而得以迅速蔓延。也就是說，即使你不認同僞科學的理念，還是可能受到它的影響。例如，有些大銀行和幾家排名財星五百大（Fortune 500）的企業，聘雇筆跡分析學家幫助他們挑選員工（Sutherland, 1992），儘管大量證據顯示筆跡學對人才選拔毫無用處（Ben-Shakhar, Bar-Hillel, Blui, Ben-Abba, & Flug, 1989; Neter & Ben-Shakhar, 1989）。從這個角度看，僞科學的筆跡學指標使雇主忽視其他更有效的選拔標準，結果導致經濟上的無效率和對個人的不公平待遇。如果你因爲筆跡中有個特別的小圓圈而沒有得到想要的工作，你會怎麼想？或者是因爲一個通靈術專家看到你的「氣場」出現一些亂流，而拒絕雇用你，你的感受如何？實際上，這確實發生在一些人身上。有些公司會付錢爲求職者進行

「通靈分析」。例如，金蘇姍（Susan King）就是一個所謂有超感官知覺、可以預見未來的人，公司付錢給她對擬聘用的人進行「預測」，幫助決定要不要雇用他們。她說她不用親自會見應聘者，從照片和應聘者的名字就可以看到她要的資料——雖然「她的雇主有時會在初選名單出來後，把她請來參與最後面試，甚至面試後的鷄尾酒會，藉以給她接近應聘者的機會」（Kershaw, 1991）。在競爭激烈的經濟時代，你要讓這種方式決定你在哪裡工作嗎？

不幸的是，這樣的例子並不少見（Shermer, 1997; Stanovich, 2004; Sternberg, 2002）。當僞科學充斥社會時，它們會以多種方式影響我們——即使你不認同這種理念。例如，警察會雇請通靈人士幫助他們調查案件（Marshall, 1980），即使研究證實這種做法並沒有增加破案的機率（Hines, 1988; Rowe, 1993）。美國廣播公司（ABC）的節目製作人請好萊塢通靈人士幫他們決定節目的內容。試想，電視是美國社會最具影響力的電訊技術，竟然由通靈人士來決定節目內容（Auletta, 1992, P. 114）。最出乎我們意料的是，在雷根總統主政時，白宮聘請一個星相學家建議「總統演講、出席、與各州州長見面、飛行旅程安排的時間，甚至還包括要討論的議題」（Johnson, 1991, p. 454）。

像星相學這樣的僞科學已經是個大產業，涉及報紙專欄、電台節目、圖書出版、網際網路、雜誌及其他傳播工具。發行量最大的是占星術雜誌，遠遠大於許多正當的科學雜誌。美國衆議院老化問題事務委員會（House Commitee on Aging）估計，全國每年被醫學騙子騙去的錢超過幾十億美元。總之，僞科學是一門相當賺錢的大生意，數以千計人的生計是靠大衆對僞科學的接受維繫的。

醫學領域如何對待僞科學這一問題，很值得心理學借鏡。因爲在抨擊僞科學，以及劃清正當的與不正當的醫學實踐界限方面，醫學界的各類專業比心理學界做得更好。讓我們看看由關節炎基金會（Arthritis Foundation）出版，曾被美國衆議院老化問題事務委員會引述的一套識別不道德藥品推銷員的指南：

1. 他是以提供「特別的」或「祕密的」處方或儀器「根治」關節炎。
2. 他刊登廣告，廣告的手法是讓取得滿意療效的「病患」講述他們的「病史」和見證經歷。
3. 他保證（或暗示）可以很快或很容易就把病治癒。
4. 他可能會說知道關節炎的病因，進而說要清洗你體內的「毒素」，「增強」你的健康。他可能還會說，不需外科手術、X光和醫師的處方藥。
5. 他可能會指責「正規醫療機構」過於謹慎，妨礙醫學進步或迫害他……但他不會讓他人用實驗和其他實徵的方式檢驗他的方法。（U.S. Congress, 1984, p. 12）

上面條列其實也可以用來作爲識別欺騙性心理治療和理論之指南。當然，指南中的第二條涉及本章的主題；但也要注意，第一和第五條說明了我們在前幾章討論過的內容之重要性：科學是公開的。除了用見證敘述作爲「證據」之外，僞科學的從業人員經常試圖逃避公開證實這一科學的標準，並指責他人有意壓制他們獲取的「知識」，這樣就有藉口帶著他們的「研究成果」直接走進媒

體，而不是透過正規的科學出版程序公諸於世。這種不正當的手段，在心理學領域中比其他領域用得更成功。因爲與其他學科相比，媒體通常看不起正規的心理學研究機制。請大家注意這一點（在第十二章我們將進一步討論這個問題）。如果是物理學的理論未經證實，媒體通常不會報導，但是未經證實的心理學理論通常會隨隨便便報導出來，這是拜僞科學經常在媒體炒作之賜，讓新聞記者以爲心理學「任憑做什麼都行」。因此，心理學知識的消費者必須意識到：如果電視和平面媒體會大肆報導心理學領域中任何稀奇古怪的說法，而不管這種說法與現有的證據是否矛盾，這一現象的受害者，最終，是一般老百姓。

各種神醫神藥的宣傳，使人對自己的病症產生過高的期望，這種期望在遭遇挫折時會給人們帶來心理損傷。在我的資料中，有關這方面最卑鄙的一個實例，是一篇關於藥店銷售藥品的文章，標題是「通靈人士顯示盲人如何用超感官知覺看東西」。人們經常因爲迷上僞科學，沒能充分利用圍繞在他們身邊的眞正知識。通靈術的支持者等於鼓勵人們在假藥上亂花錢，並讓患者忽視眞正有可能幫助他們的傳統「非通靈」醫學治療（Angell & Kassirer, 1998）。記者布倫尼曼（Richard Brenneman, 1990）描述了通靈外科醫師如何騙光癌症患者的錢，並延誤了本可透過合理醫學手段進行的治療。類似地，在一個有關醫療保健欺騙行爲的市民大會上，密西根州反健康欺騙委員會（Michigan Council Against Health Fraud）的主席展示了一則虛假癌症治療的宣傳册子，上面宣稱「本產品不能與其他癌症治療方法同時使用」（「If It Sounds」，1990）。類似案例造成的損害是無法估算的。

CH.12

醫師愈來愈關注醫學騙局在網際網路上的蔓延，以及人類健康爲此付出的代價。科普斯（Max Coppes）醫師曾經因氣憤而致信《新英格蘭醫學期刊》，警告大家醫學騙子對人類造成的眞正傷害（Scott, 1999）。他講了一個九歲女孩的故事。癌症手術後，如果她接受化療，有50%的機率再活三年。然而她的父母發現一種未經驗證有效的治療法——用鯊魚軟骨組織治病。他們採用了這種治療法，結果那個小女孩只活了四個月。

因此，任由僞科學蔓延的代價是巨大的。僞科學蔓延的主要原因就在於，人們混淆了哪類證據可以、哪類不可以用來增加某個說法的可信度。由於見證叙述能很方便地爲幾乎所有理論及說法提供支持，並透過這些支持造成巨大的影響，見證型證據就爲僞科學的發展和獲得信任敞開了一扇大門。對心理學知識的消費者來說，除了小心提防外，沒有其他更好的辦法。在接下來幾章，我們會討論要證實一個理論或說法，到底需要哪一類型的證據。

摘要

個案研究和見證叙述在心理學（及其他社會科學）的研究初期是有用的。因爲那時我們發現可以進行研究的現象和變數是至關重要的。雖然個案研究在早期的、理論形成前的科學研究階段是有用的，但是在研究的後期——對提出的理論進行特定檢驗時——它就變得毫無用處了。在心理學中，

個案研究和見證敘述對理論檢驗沒有價值的一個主要原因是，有安慰劑效應作祟。安慰劑效應是指人們有報告任何治療對他們都有幫助的傾向，不管是不是眞的有療效。安慰劑效應的存在，使我們不可以採用療效見證作爲支持一個心理（或醫學）治療有效性的證據。原因是安慰劑效應保證：不管什麼治療方法，都有可能獲得有關它的療效的見證資料。

儘管見證證據對理論檢驗毫無用處，心理學研究還是發現：這類證據由於鮮活性效應而被人們格外看重。鮮活性效應是指：人們過分看重更鮮活和更容易從記憶中提取的證據。對大多數人來說，見證證據就是一種極具鮮活性的資訊。其後果是：人們在驗證某一心理學的理論說法時，往往過分依賴這類證據。實際上，見證敘述和個案研究型證據，根本不能用來支持任何理論說法。

第五章 相關和因果

——用烤麵包機避孕?

幾年前,台灣做過一個大規模的調查,研究哪些因素與人們是否使用任何避孕方法有關。由一些社會學家和醫學工作者組成的龐大研究隊伍,廣泛蒐集了諸多行爲和環境變數的資料。研究者想知道什麼變數最能有效預測人們是否使用任何方法避孕。蒐集資料後,他們發現人們是否使用任何避孕方法與一個變數的相關最高:家中擁有家用電器(如烤麵包機、風扇等)的數量(Li, 1975)。

這個研究結果,恐怕不會讓你想用在各高中免費發放烤麵包機,以解決靑少年的懷孕問題吧!但是,你爲什麼不會這樣想呢?既然小家電和是否使用任何避孕方法之間似乎存在著密切的關係,而這個變數又是所有當時測量的衆多變數中最具預測性的一個,你應該沒有理由不這樣想。我希望你對這個問題的回答能是:在這個研究結果中,是兩個變數關係的**本質**而非其強度與研究者想理解的問題相關。試想一下,如果我們舉辦一個「在高中免費發放烤麵包機」的活動,人們會怎麼想?大家會以爲是烤麵包機導致人們避孕!這不是很荒謬嗎?在這個顯而易見的例子中,我們很容易看

出這兩個變數之間或許有關聯，但決不可能存在因果關係。

在這個例子中，我們不難猜到，避孕和擁有小家電數量之間的關聯，可能是因爲這兩個變數同時與某個第三變數都有相關所產生的結果。例如教育水準就是一個可能的中介變數，我們知道，一個人的教育水準與其是否使用任何避孕方法，以及與其社經地位都有相關，而社經地位高的家庭，往往擁有較多的家用電器。當然，兩者之間可能還有其他更多的變數起作用，在此僅想說明，無論烤麵包機的數量與避孕有多大的相關，它們之間也不會意味著有因果關係。

這個例子應該很容易讓我們理解本章要談的基本原則：有相關並不表示一定有因果關係。在本章中，將討論讓我們無法做出因果推論的兩大問題：第三變數問題與方向性問題。在討論完這兩個問題之後，我們還會討論第三變數這一問題經常是由選擇偏差（selection bias）引起的。

相關作爲研究證據的局限性有時很難察覺。當我們以爲一定存在因果關係時，當我們心存偏見時，或當我們一心只想尋找證據支持自己的理論時，很容易會將相關看成因果關係。皮爾遜（Karl Pearson）的個案提供一個清楚且諷刺的例子。

第三變數問題：古德伯格與癩皮病

二十世紀初，數以千計的美國南部人遭受一種稱爲癩皮病（pellagra）的皮膚病侵襲，每年約有

十萬人喪命。症狀有頭昏眼花、昏睡無力、傷口不癒、嘔吐、嚴重腹瀉。這種病被認爲是某種「來歷不明」的微生物感染所致（Chase, 1977, p. 205），當時有研究資料顯示，此一病症與衛生條件有關。這一結果，連全國癩皮病研究學會（National Association for the Study of Pellagra）的醫師都認爲很有道理，因爲在南卡羅萊納州的斯巴坦堡（Spartanburg）市內，沒有受癩皮病干擾的人住家似乎都有自來水及污水管道系統。也就是說，衛生條件較好。這種相關恰好驗證了當時對傳染病病理的一個理論：由於衛生條件較差，污水處理不當，癩皮病透過感染者的糞便不斷傳播。

古德伯格（Joseph Goldberger）醫師對這種解釋不以爲然，他在美國公共衛生部部長的指示下對癩皮病進行好幾個調查研究。古德伯格認爲癩皮病是營養不良引起的，因爲美國南方地區普遍比較貧窮，人們主要依靠穀物、燕麥和玉米粥等高碳水化合物的食物維生，而少吃肉、蛋、奶等蛋白質含量較高的食物。古德伯格認爲，污濁的環境和癩皮病之間的相關並非因果關係（就像烤麵包機與節育行爲的例子），之所以會出現相關，是因爲有下水設施的家庭通常經濟條件也比較好，而這種經濟差別也反映在他們日常的飲食中，例如是不是食用較多的動物蛋白質。

但是，好像有些不對勁！爲什麼我們要相信古德伯格的營養缺乏理論，而不相信前面的糞便傳播理論呢？畢竟，兩派人馬都是坐在那裡根據相關資料推論什麼才是造成癩皮病的原因。爲什麼醫學會的醫師不能說古德伯格的相關同樣是在誤導大家？古德伯格又有什麼理由排除糞便處理不當這一理論假設呢？古德伯格之所以理直氣壯，是因爲他做了一件我剛才忘了說的事：他曾親自吃下癩

皮病患者的糞便！

◎為什麼古德伯格的證據比較好

支持古德伯格之理論的證據，不僅包括觀察變數間相關係數所做的推斷，還包括另一類研究證據：對關鍵變數進行操控之後所蒐集的資料（有關操縱控制，下一章再進一步討論）。這一方法通常涉及佈置一個在自然情況下極少出現的特殊場景。為了製造這樣的特殊場景，如果我們說古德伯格想出一個相當不自然的方法，決非言過其實。

為了證實癩皮病不會傳染，不是透過患者的體液傳播，古德伯格在自己身上注射患者的血液，還將患者喉中和鼻中的分泌物放入自己口中。除此之外：

> 最後，他選出兩位癩皮症病患，一個有皮癬症狀，一個有腹瀉症狀。他從病患身上撥下魚鱗狀角質皮混入四毫升該病患的尿液，並加入四毫升他們的糞便，然後與四小撮麵粉揉在一起做成小藥丸。他自己及幾個助手，還有他的妻子，志願吞服下去。（Bronfenbrenner & Mahoney, 1975, p. 11)

古德伯格和其他志願吞食的人都沒有染上癩皮病。這就說明，他已經佈下這種病透過感染傳播的所有可能的途徑，結果什麼事也沒發生。

古德伯格設想並佈置了幾套操作程序，以控制別人認爲是癩皮病的「因果機制」之變數，從而證明此一機制對該疾病的無效性。接下來，他的任務是證明自己的因果機制有效。他從密西西比州監獄農場找到兩批無癩皮病並志願參加實驗的犯人，其中一批配給高碳水化合物／低蛋白質的食物，這是他預期會導致癩皮病的原因；另一批人則配給相對營養均衡的食物。五個月內，食用含高碳水化合物／低蛋白質食物的犯人果然爲癩皮病所肆虐，另一批人則安然無恙。儘管後來古德伯格的觀點，遭到一些因政治目的而企圖否定美國存在貧窮人口（到吃不上蛋白質食品的地步）現象的人強烈反對，但經過一段長時間的打壓，由於他的理論確實比其他人的得到更多實徵資料的支持，最終還是得到認可。

癩皮病的這段歷史告訴我們，如果依據相關研究的基礎制定社會和經濟政策，必將使人類付出慘痛的代價。但這並不是說我們最好永遠不要使用相關資料作爲研究證據；恰恰相反，在許多情況下（例如只想預測、不探究原因時），它正是我們需要依靠的證據（見第八章），科學家經常必須在沒有充分資訊及知識的情況下解決難題。重要的是，我們在運用相關資料時應該小心謹慎。像癩皮病－污水案例在心理學領域中也經常發生，這就是**第三變數的問題**。癩皮病－污水案例中的兩個變數的相關，並不足以說明它們之間有某種直接的因果關係；這一事實說明的，很可能是與這兩變數都相關的第三個變數造成兩者的相關，在該例中，這第三個變數是飲食營養。癩皮病和污水的這種相關通常稱爲**虛假相關**（spurious correlations），因爲兩者出現的關聯，並不是由於兩個變數之間存在因果

關係，而是兩變數均與第三變數相關聯所致。

再來看一個現實生活中的例子。公私立學校教學品質的比較一直是人們爭論的話題。在這個爭論中所下的一些結論，相當鮮活地展示了從相關資料推斷爲因果關係的弊端。私立學校比公立學校好還是壞，應該是用實徵研究解決的問題，可以透過社會科學中的調查法分辨眞僞。但並不表示這是一個簡單的問題，只是說明它是一個科學問題，有解決的可能性。所有倡導私立學校比較好的人對這一點心知肚明，因爲在他們維護自己的觀點時，常常用這樣一個實徵研究的結果：私立學校學生的成績比公立學校好。這一事實無可爭辯，各類研究和教育統計資料在這方面都得到很明顯也很穩定的結論。但是問題在於，僅用這些學生的成績，是否眞的就可以讓我們下結論說：私立學校的教育本身導致他們學生成績比較好？

教育效果測驗的結果是與許多因素相關聯的，而這些因素彼此之間又都互有關聯。要評估公立與私立學校的好壞，我們不能僅用學校類型與學生成績之相關值這個指標，例如學生成績與家庭環境因素、父母的教育程度、父母的職業、家庭的社經地位、家中藏書數量等都有關係。這些因素與孩子能否進入私立學校的可能性又有相關。因此，家庭環境作爲一個潛在的第三變數，可能是導致學生成績與學校類型相關的原因。總之，學生成績可能與學校品質無關，家庭經濟條件好、有良好教養的孩子，本來就更有可能就讀私立學校，這類學生即使在公立學校上學，學業成績也會很好。

幸運的是，現在有許多像**多元迴歸**（multiple regression）、**路徑分析**（path analysis）這類複雜的相

關統計方法（順便提一下，統計學的發展許多是心理學家的功勞），可以幫我們釐清這些問題。這些統計方法可以計算出兩個變數在排除其他變數干擾之後，比較「純淨」的相關值。杜克大學（Duke University）的佩奇和基思（Ellis Page & Timothy Keith, 1981）曾用這種複雜的相關技術，分析在美國全國教育統計中心（National Center for Educational statistics）的幫助下蒐集的高中生教育統計資料。他們發現，排除一些影響學生家庭環境和一般性智力因素之後，學生學業成績與學校類型之間不存在任何相關，此一結果也得到其他研究者的確認（Berliner & Biddle, 1995; Wolfle, 1987）。

因此，提倡開辦私立學校以加速教育發展，和用烤麵包機控制生育，看來是沒什麼兩樣。學生成績與就讀學校相關不是因爲它們有直接的因果關係，而是因爲家庭環境及智力水準的差別罷了。

複雜相關統計方法可以幫助我們排除第三變數的影響，但並不表示這個排除過程一定會削弱原有相關的強度。有時候，當我們排除第三變數之後，兩變數之間相關的強度仍保持不變。這時，這一相關就告訴我們許多東西了。很可能，兩者的相關並非由第三變數造成，而是的確存在。當然，也有可能是還有一些沒有想到的，或無法排除的其他變數造成虛假相關。

湯瑪斯、亞歷山大和埃克蘭德（G. Thomas, K. Alexander, and B. Eckland, 1979）的研究，也提供了一個很好的例子。他們的調查發現：高中生能否上大學與家庭的社經地位相關。這個重要發現，無疑擊中了美國這樣一個以能力爲價值核心的社會之要害，因而引起很大的反響。因爲這一研究結果不正是在暗示，一個人一生會不會成功不是取決於能力及努力，而是取決於他的社經地位嗎？然而，

在下這個結論之前，也許應該先考慮一些可能用以解釋此一資料的其他假設；也就是說，上大學與社經地位之間的相關很可能只是一種假象。很容易想到的一個對兩者都有影響的可能因素是學生的學業能力，它與讀大學及社經地位兩者都有關係。如果將這個因素考慮進去，那麼原來那兩個變數之間的相關或許就不存在了。但當調查人員處理過這個第三變數、重新對兩者進行統計分析時，卻發現即便排除此一第三變數，上大學與社經地位之間仍然存有顯著相關。因此，看來高經濟收入家庭的孩子更容易考上大學，並不完全是由於學業能力比較強。正如前面所說，這個結果當然並不能排除其他第三變數造成兩者相關的可能性，但是能夠用重新分析以排除學業能力對兩者相關的影響，本身無疑在理論及實踐方面也具有明顯的重大意義。

安德森與安德森（C. A. Anderson and K. B. Anderson, 1996）就曾透過檢驗其他假設能否解釋同一組研究資料的方法，驗證一個關於暴力具有地區差異的理論。他們使用的方法就是偏相關（partial correlation）。有研究資料顯示美國南部的暴力犯罪率高於北部各州。安德森與安德森（1996）對這個**高溫假設**（heat hypo-thesis）進行驗證，這一假設認爲南部地區所以有較高的暴力事件，是因爲「令人不舒適的高溫導致人們的攻擊性動機和（有時候）攻擊行爲的增多」（p. 740）。找到一個城市平均溫度與其犯罪率之間有相關並不難，許多人都報告了這一結果。然而，當他們發現即使將失業率、人均收入、貧窮率、教育狀況、人口數量、人口平均年齡及其他幾個可能是第三變數的因素都加以控制之後，溫度與暴力犯罪之間的相關仍然顯著。這就使得高溫假設理論的可信度大大提高了。

方向性問題

如果我們可以透過操控變數順理成章推斷因果關係，就沒有理由像上述那樣拐彎抹角地靠分析相關係數來推論。但讓人苦惱的是，一旦涉及心理學的問題，我們經常必須靠對相關係數的進一步分析，而且在心理學知識對解決社會問題益發重要的今天，這種依賴相關分析的趨勢益發嚴重。教育心理學裡一個廣為人知的例子，很恰當地說明了這一點。

人類對閱讀進行科學研究大概始於一百年前。從那時起，研究者就發現眼動形式（eye movement patterns）與閱讀能力相關，閱讀能力差的人在閱讀時眼動沒有規律，並經常回掃（從右到左）和凝視（停下不動）。基於這一相關，一些教育學家認為缺乏眼動技巧是導致閱讀困難的原因。於是，許多眼動訓練專案開始展開，並且很快地應用於小學中。在還沒有弄清楚這兩個有相關的變數到底哪個是因、哪個是果的情況下，這些教學專案早已啓動施用了。

今天我們已經很清楚，人們從這一個相關推論出的由眼動到閱讀能力的因果關係，恰恰與實際情況相反。無規則的眼動不會導致閱讀困難（Olson & Forsberg, 1993; Rayner, 1998）；而是閱讀速度慢、理解困難導致無規則的眼動！小孩子在識字水準和理解力提高的過程中，眼動會逐步趨於有規則；相反地，訓練孩子的眼動不會提高他們的閱讀水準。

最近十幾年，研究者已經清楚地發現，詞語編碼及語音加工時出現的語言問題，才是閱讀困難

的根源所在（Rayner, Foorman, Perfetti, Pesetsky, & Seidenberg, 2001, 2002; Shaywitz, 1996; Stanovich, 2000），而很少發現不良的眼動形式導致閱讀能力喪失的例子（Olson & Forsberg, 1993; Share & Stanovich, 1995）。但是如果你現在到中等或以上規模的學校地下儲藏室仔細找一找，一定會發現一個蓋滿灰塵的眼動訓練儀，如今，這些儀器已成爲浪費公物的樣版。這全是由於人們習慣把一個相關當作因果關係造成的惡果。

另一個有些相似的例子第一章中討論過。在教育社會工作領域裡，曾一度流行學業成績低、濫用毒品、少女懷孕及其他許多問題行爲是低自尊造成的（Dawes, 1994; Kahne, 1996）。這個因果關係的方向似乎理所當然：低自尊導致問題行爲，高自尊導致優秀的學業成績及其他成就。因此在這一因果推論的激發下，各類提升自尊的訓練專案應運而生。但這裡的問題與眼動的例子一樣：這個因果關係的結論，僅僅是從兩個變數的相關關係中推衍出來的（更何況在這裡相關值非常小，請參見 Kahne, 1996）。如今證實，即便是自尊與學業成績存在因果關係，方向卻極可能與前面的推論正相反：是在學校（以及生活中其他方面）的優秀表現導致高自尊（Baumeister, 1999; Ruggiero, 2000; Stout, 2000）。

我們前面的討論，主要是圍繞著兩變數的相關常常涉及的兩類問題：第一類叫做**方向性問題**，眼動和自尊的例子可以說明，當變數A與B相關時，我們在下結論這一相關是由A導致B之前，必須想到這種因果關係的方向很可能相反；也就是說，很可能是由於B的變化而引起A的變化。第二類是第三變數問題，癩皮病（和烤麵包機與生育控制，以及私立學校與學業成績）的例子可以說明，兩變

數之間的相關，或許不能告訴我們一方是另一方的因或果，很可能只是兩個變數均與第三變數有相關而導致的。

選擇偏差

在一些情況下，虛假相關存在的可能性是很高的——當**選擇偏差**出現的時候。所謂選擇偏差是指：某些人與某些環境存在著一種關係，是由於具有某些特定生理、行爲、心理特徵的人會選擇生活在具有某一特定的環境居住。因此，選擇偏差造成環境特徵與行爲及生理特徵之間的虛假相關。

讓我們透過一個例子看看選擇偏差如何產生虛假相關。請你很快地告訴我們美國哪一州，由呼吸道疾病導致死亡的人數最多。答案當然是亞歷桑那州！爲什麼？亞歷桑那州的空氣不是已經淨化得很好嗎？難道說洛杉磯的廢氣濃煙會擴散到那麼遠嗎？鳳凰城的郊區環境已經變得那麼差了嗎？不可能，絕對不可能！別急，讓我們好好想想：或許亞歷桑那州的空氣的確很**清新**，或許只是有呼吸道疾病的人喜歡移居到那裡，然後就在那裡去世。呀！這樣想就對了。這讓我們注意到，如果一不留意，就很可能受呼吸道疾病與死亡人數的相關誤導，以爲是亞歷桑那州的空氣致命的。

然而，選擇偏差有時很難發現，它們經常受到忽略，特別是像自尊的例子那樣，當我們心中早已存在一個特定方向的因果關聯的想法時。這時，即使是再聰明的頭腦，誘人的相關係數和心中的

預設偏見也會聯合起來矇騙我們。下面是一些具體的事例。

有關美國教育品質這一問題，過去二十年人們一直爭論不休。在這一問題的辯論中，我們可以看到考慮選擇因素的重要性。當時，大量的統計資料曾一股腦地湧向大衆，卻沒有提醒、警告他們要避免以相關資料推論因果關係，因爲相關資料內含大量具有誤導性的選擇偏差。

在整個爭論過程中，許多帶有政治目的的教育評論家，一再試圖提供有關教育品質與教師工資水準及班級大小規模沒有關係的證據，儘管已經有許多的研究表明，後兩個變數對前者都很重要（Ehren-berg, Brewer, Gamoran, & Williams, 2001; Finn & Achilles, 1999）。他們提到的證據中，有一個是對五十州的學生之大學入學性向測驗（Scholastic Aptitude Test, SAT）結果所做的分析。資料顯示，測驗的平均分數的確與教師工資和教育支出的相關很小。即使有相關也與人們期望的方向相反。一些教師平均工資高的州，學生測驗成績的平均分很低；而有些教師平均工資低的州，學生的測驗得分卻很高。但是，在這個研究中，只有那些想考大學的高中生才參與這次測驗，這就使我們看到選擇偏差導致虛假相關有多容易。

從測驗中我們看到，密西西比州的學生SAT的得分高於加州的（Powell & Steelman, 1996; Taube & Linden, 1989），而且差異相當大，密州比加州的平均分數高出一百分之多。由於密西西比州的教師工資是全國最低的，所以爲那些極力要求削減教師工資的保守派人士提供一個慶祝的理由。但是，密西西比州的學校是否眞的比加州好？再者，是不是前者的整體教育狀況都比後者好？當然不是。

幾乎所有其他的客觀指標都顯示加州的學校更勝一籌（Powell & Steelman, 1996）。果眞如此，又如何解釋SAT測驗的結果呢？

答案就在於選擇因素。不同於許多要求全州教育系統的所有學生都要參加的標準化測驗，SAT測驗不是所有高中生都要參加，因而存在選擇偏差（Powell & Steelman, 1996; Taube & Linden, 1989; Wainer, 1989）。只有那些認爲自己有希望進入大學的學生才會參加這項測驗。這個因素是導致各州測驗的平均分數存在差異的部分原因，同樣也解釋了爲什麼一些教育實力最強的州，在SAT測驗中平均分偏低的現象。

選擇因素以兩種途徑作用於各州的SAT分數。首先，一些州立大學入學的要求是看美國大學入學考試（American College Testing, ACT）的分數，而不是SAT。因此，這些州參加SAT考試的都是想到其他州讀書的學生，而能夠去外地讀書的學生往往家庭條件比較好，學業能力也高於該州平均水準。這就是爲什麼會出現密西西比州與加州的差異情況：在密西西比州只有4％的高中生參加SAT測驗，而加州則是47％（Powell & Steelman, 1996）。

第二個選擇因素則微妙得多。在一些教育系統發達的州，許多學生喜歡讀完高中以後繼續讀大學深造，所以在這些州中，很高比例的學生——包括相當一部分能力稍差的學生——也都參加了SAT測驗。而在一些輟學率高、教育品質較差的州，只有較少的學生想讀大學，因此，這些州最終參加SAT的學生僅僅是最有能力上大學的那群人。這樣一來，當然是這些州SAT的平均分要高

於那些報考比例大的州。

錯用SAT分數讓我們看到，當大衆缺乏本書所說的基本方法和統計思想訓練時，要想糾正誤用統計是何等困難。一九八三年本書的初版中，我就以SAT分數爲例解釋選擇偏差。十多年後的一九九四年，在第四版中我又討論了一位印第安那州教授鮑爾（Brian Powell, 1993）的一篇關於分析保守派政治專欄作家威爾的文章。如你所料，威爾在文章中聲稱由於高SAT分數的州教育投資並不高，所以他反對在教育中撥款太多。然而，鮑爾指出，那幾個單獨列出的所謂分數特別高的州：愛荷華州、北達科他州、南達科他州、猶他州和明尼蘇達州，其SAT測驗的參加率分別僅只有5%、6%、7%、4%和10%；相反地，全美國有超過40%的高中畢業生參加了SAT測試。因爲在這些州中，公立學校所需的入學測試爲ACT測試，只有那些計畫到其他州，而且「通常是到聲譽很好的私立學校」讀書的學生才參加SAT測驗（Powell, 1993, p. 352）。相反地，在威爾點名爲SAT得分低而教育消費高的新澤西州中，卻有76%的高中畢業生參加SAT測驗。顯然，北達科他和南達科他州參加SAT測驗的學生，與有四分之三的學生參加這項測試的新澤西州的學生相比，是更是優秀的一群。

在《教育研究者》（*Educational Researcher*）雜誌上，心理測量學家韋納（Howard Wainer）曾分析一九九三年六月二十二日發表在《華爾街日報》（*Wall Street Journal*）的一篇文章。這篇文章介紹一個由美國傳統基金會（Heritage Foundation）所做的研究。該基金會是一個帶有較強烈社會政治理念色彩的

智囊團，反對在教育上投入太多的經費，他們的證據是，想必你已經猜到了，SAT分數較低的州在教育中的投資反而較多。所幸，韋納的文章不僅解釋是選擇偏差導致這一結果，還讓我們看到，當把一個測驗——全國教育進展評估（National Assessment for Educational Progress, NAEP）——施於具有代表性的樣本，而不是順著自主選擇所得到的樣本時，這種關係的因果方向會反過來：在教育中投資愈多的州，平均得分愈高。

運用前面提到的偏相關分析法，鮑爾和史提爾曼（B. Powell and L. C. Steelman, 1996）也證實了這種關聯。他們發現，當用統計學的方法將這些州參加測試的學生修正到比例相同時，每州在每個學生身上增加一千美元的教育投資，與其SAT測驗平均分數增加十五分有相關性。然而，儘管有大量證據表明如果不作統計修正，選擇偏差將使得各州之間的SAT分數不具可比較性，但媒體和政客爲了他們的政治目的，仍繼續使用這些未經修正的數字誤導大衆。

臨床心理學的一個事例可以表明選擇偏差問題有多微妙和「邪惡」。研究資料經常表明，對各類成癮問題，如：肥胖、吸煙、吸毒等，做過心理治療的人的治癒率低於那些沒做過的（Rzewnicki & Forgays, 1987; Schachter, 1982）。其原因，你會很高興知道，並不是心理治療使成癮行爲更加難以改變，而是在於那些尋求心理治療的人，其成癮問題比較不容易察覺，而且無法用自療來改善病情。

韋納（1999）講述了二次世界大戰中的一個故事，也讓我們想到選擇偏差邪惡的一面。一個飛機分析師想透過分析出擊後返航的飛機被子彈擊中的彈孔分佈，以確定在飛機上安放加固防彈層的

位置。經過分析後他認為，應該將防彈層安置在返航機上沒有被子彈襲擊的位置。他的理由是，子彈應該是均勻地襲擊飛機的各個部位。所以，如果一架飛機能返回，就表示這架飛機被子彈擊中的地方必定不會對該飛機造成致命損傷。那些沒有彈孔的地方，看來都是要害，因為被擊中該處的飛機都沒能返航。因此，防彈裝置應該安裝在返回飛機中沒有被子彈擊中的部位！

簡而言之，本章推薦給消費者的原則很簡單：注意選擇偏差的可能性，並應避免僅僅從相關資料推論因果關係。當然，我們的確可以透過某些複雜的相關設計做出有限的因果推論，某些研究問題由於結構的特殊性，也可以用相關資料作為檢驗幾種因果假設的基礎。同樣地，相關證據也有助於表明眾多研究證據已經有趨於支持一個假設的現象（見第八章）。儘管如此，對於心理學知識的使用者來說，寧可犯太挑剔的錯誤，也不要被相關資料欺騙，誤認為兩個變數有因果關係。

摘要

本章的中心議題是想告訴大家，兩個變數之間僅僅存在相關，並不能保證一個變數的改變會導致另一個改變——相關並不意味著因果關係。在解釋相關關係時，我們討論了兩個問題。首先，在第三變數問題上，兩個變數相關並不意味著它們之間有直接的因果關係，或許是由於一個尚未被測量的、與它們倆都有關聯的第三變數所導致。實際上，如果測量了潛在的第三變數，就能夠用相關

統計法——如偏相關（第八章會再討論）——評估第三變數是否決定了二者之間的相關。第二個解釋相關的困難，在於存在方向性問題：即使兩變數間確實存在直接因果關係，到底哪個是因、哪個是果，也無法僅從相關中反映出來。

選擇偏差是行為科學中許多虛假相關產生的原因。所謂選擇偏差是指：人們總是或多或少人為地選擇自己的居住環境，因此造成某些行為特徵與某些環境變數之間有相關。如本章古德伯格的例子所展示的，以及下兩章會詳述的：排除選擇偏差的唯一方法，就是在嚴格操控各變數的情況下進行眞正的實驗研究。

第六章　一切盡在掌控中

——聰明漢斯的故事

本章一開始，讓我們先做個測驗。噢！別擔心，不是測驗你前幾章讀得好不好。其實這個小測驗非常簡單，它與我們平時見過的一些物體移動現象有關。請準備好一張紙，開始了。第一題，想像一個人拿一根繩子在頭頂上繞圈，繩子的遠端繫著一顆球。從人頭頂的正上方，我們可以看到這顆球運行的軌跡，是個圓圈。請將這個圈畫在紙上，並在這個圈的一處畫上一個點。然後，在這個點和圓心之間畫一條直線，直線代表繩子，點代表在特定時刻的那顆球。假設在某一旋轉瞬間，繩子被剪斷了，你的第一個測驗任務就是用筆畫出這個球飛出後的運行軌跡。

下一個測驗任務是，想像你是個轟炸機的飛行員，現在正以每小時五百哩的速度在二萬呎的高空衝向目標，爲簡單起見，假設沒有空氣阻力，那麼你要在哪個點上投放炸彈才能擊中目標：到達目標前？目標的垂直上空？還是飛越目標之後？請說出你選擇的投放點與目標間的具體距離。

最後，想像你正在進行槍與肩同高的步槍射擊運動，假設沒有空氣阻力，並且槍與地面平行。如果一顆子彈從與槍相同的高度墜落，到地面需一．五秒。假設你現在由槍管中射出一發子彈，起始速度為每秒二千呎，它需要多長時間才能落到地面？

好了，測驗完畢！答案？噢！對了，答案我們會在後面公佈。但是首先，為了了解對運行物體知覺的準確性，我們需要更全面地了解科學家的實驗方法背後的思考邏輯。在這一章，我們將探討實驗控制和操弄（manipulation）的一些原理。

史諾與霍亂

在前一章我們講到，古德伯格對癩皮病的研究，部分靠的是他認為這種病是不會傳染的靈感。然而在七十年前，史諾（John Snow）在研究霍亂病因時，則將病因壓在相反的猜想上，最後也成功了（Goldstein & Goldstein, 1978）。早在一八五〇年代的倫敦，人們對不斷爆發的霍亂提出許多理論，而且彼此爭論不休。很多醫師認為霍亂病患呼出的氣體會導致感染，這就是所謂的**穢氣理論**。史諾的想法卻是：疾病是透過被病患排泄物污染的供水系統傳播出去的。

史諾決定檢驗自己的想法。幸運的是，在倫敦有很多水源，不同的水源供給不同的地區。如果他的構想是對的，霍亂的發生率應該因水源受污染程度的不同而有差別。但是史諾也想到，這樣的

比較可能會出現嚴重的選擇偏差（請回想一下第五章的討論），由於倫敦各區貧富差別很大，因此，水源與各地區的患病率之間的任何相關，都可能是其他影響健康的、與該地區經濟發展水準相關的變數——如飲食、壓力、職業安全及衣食住行的品質等——導致的。總之，史諾得到的相關是虛假相關的可能性，和第五章講述的癩皮病與污水系統的例子一樣高。然而，史諾非常機靈地注意到這個問題，而且利用當時的一個特殊情境解決了。

在倫敦某區，恰巧有兩家自來水公司提供飲水給當地居民，而且是隨機的供水；也就是說，即便是緊鄰的兩個家庭，可能是不同的自來水公司供水。史諾並且發現，這個地區內，人們的社會經濟狀況基本上非常接近。然而，即使在這麼得天獨厚的條件下，如果兩家自來水公司的水都同時受到污染，史諾還是無法驗證其假設。幸好，史諾福星高照，兩家公司的水並未同時受到污染。

在一次霍亂流行過後，蘭博斯自來水公司（Lambeth Company）為了免受倫敦污水系統的污染，把公司遷到泰晤士河上游。而南沃克及沃克斯霍爾公司（Southwark and Vauxhall Company）卻仍待在該河下游，這就使蘭博斯公司的水受污染的可能性比南沃克及沃克斯霍爾公司的小很多。史諾運用化學檢驗證實了這一點。下面是飲用不同供水公司所提供的水之家庭的霍亂死亡率：蘭博斯為每一萬家中有三十七人死亡，而南沃克及沃克斯霍爾為每一萬家三百一十五人。

本章我們將討論在史諾和古德伯格的故事中，到底隱含著什麼科學思考的邏輯。這對我們理解科學家到底在做什麼非常重要。如果不能理解這種邏輯，他們所做的事就很可能被看成神祕的、古

怪的、甚至荒謬可笑的。

比較、控制和操弄

儘管市面上已經有很多厚重的書講授科學研究方法，但是對於從未做過實驗的門外漢來說，這些書大概都是不必要的艱難，因爲他們只想知道個大概，並不想被實驗設計的所有複雜細節絆住。科學思考最重要的幾個特點其實非常容易掌握，那就是比較、控制和操弄。爲了對一個現象有更基本的了解，一位科學家會對各種情況進行比較。如果沒有比較，我們看到的都是一些孤立的觀察或事件。就像在第四章中看到的那樣，對這類觀察的解釋往往是非常模糊的。

透過在不同條件下比較資料的異同，科學家可以推翻對資料的某些解釋從而確定另一些。他們建構比較，以排除一些同樣可能進一步成爲解釋其研究結果的其他說法或理論。也就是說，我們要在一個實驗設計中盡可能地排除不正確的可能解釋。要做到這一點，他們或者透過直接控制實驗條件，或是透過觀察允許進行比較各種可能解釋的自然情境。

後一種途徑在霍亂這個例子中得到很好的體現。那兩家自來水公司並不是史諾隨便找的，他意識到，雖然有許多地區都是由不同的自來水公司供水，但是這些地區在與健康相關的許多社會經濟指標上，也會存在很大的差異。因此，即使觀察到霍亂在不同地區的不同發病率，也難逃面臨「許

多不同解釋可以同時存在」的難題。史諾非常清楚，只有排除無關的解釋才能推動科學發展（請回想第二章討論的可證偽性），所以他努力尋找，並且找到一個比較情境，在此情境中，許多認爲霍亂疫情與社經水準相關的大量其他可能解釋都排除了。

史諾能找到這樣一個消除其他解釋的自然情境，是非常幸運的。這種在自然情況下產生的比較條件不多見。但如果因此以爲，科學家都是在那裡等待這種情境的出現，那就太荒謬了。事實上，大部分科學家都嘗試著重構我們的世界，從而可以用之分辨各種可能的假設，看看哪些假設是對的，哪些是錯的。要做到這一點，就必須操弄那個可能是原因的變數（如史諾例子中的供水污染），並觀察，當保持其他變數恆定時，是否會產生不同的結果（霍亂發病率）。

因此，一個好的實驗設計應該是由科學家操弄他感興趣的變數，並控制其他可能造成影響的無關變數。需要注意的是，史諾並不是這樣做，因爲他沒有辦法操弄水源污染的程度。但是他幸運地找到一個自然情境，允許他在其他變數（主要是與社會經濟水準有關的變數）受到控制的情況下，觀察水源污染的變化。

讓我們拿史諾的兩組受試再思考一下：一組由蘭博斯公司供水，一組由南沃克及沃克斯霍爾公司供水。由於處在同一地區，可能保證了兩組受試者的社會地位水準幾乎相同。然而，史諾用的這一型研究設計有個缺點：它是由受試者決定自己屬於哪一組，因爲他們早在幾年前已與兩家公司簽訂了供水合同。我們因此還必須考慮是什麼原因促使他們與該家公司簽約，而不是另外一家。是由

於這家的價錢比較便宜，還是廣告宣稱他們的水質好、具有療效？這些我們並不知道。關鍵問題是，選擇那一家宣傳水質有療效的公司的家庭，會不會在與健康相關的其他因素上也與其他家庭不同，而或許這些因素才是低發病率的眞正原因。這是一種可能性。

史諾這一類型的設計也不能排除比較隱晦的虛假相關，這些虛假相關可不像前述那些與社會經濟因素掛鉤的那樣容易辨認。這也是爲什麼科學家寧可直接操弄他們感興趣的變數。當操弄與一個叫做**隨機分派**（random assignment）的程序（在這一程序中，受試者不能決定自己加入哪一條件的實驗組，而是隨機地分派到其中一組）結合時，科學家就可以排除那些歸因於受試者本身不同特徵的解釋了。當然，這裡並不是說史諾的方法沒有價值，只是科學家會更喜歡直接操弄實驗變數，因爲這樣推論才更具說服力。隨機分派可以確保實驗條件中的受試者在所有其他變數上相當一致，隨著樣本人數的增加，它還可以抹平一些偶然因素。因爲這種分派方式是完全由沒有偏私的隨機抽樣工具（例如一枚硬幣、亂數表，等等）來選取樣本，沒有經過人爲的挑選。

至於古德伯格，他的確操弄了研究假設中的原因變數。在一連串研究中，他不僅發現並記錄與癩皮病有關的變數，還直接控制研究中的其他兩個變數。請回想一下，他安排低蛋白質飲食的囚犯組以誘發癩皮病，以及吞食癩皮病患者排泄物的志願組，其中包括他和妻子。這樣，古德伯格不僅觀察自然存在的相關，還創設了特殊條件組，從而排除更多不同的可能解釋，也因此他的推論就比史諾的更有說服力。這就是科學家爲什麼試圖操弄變數並保持其他變數恆定：爲了排除可能替代自

己的解釋之所有其他解釋。

醫學中不乏對實驗缺少控制而導致錯誤結論的例子。羅斯和尼斯貝特（L. Ross and R. E. Nisbett, 1991）提到一個發生在一九六〇年代中期的案例。門腔靜脈分流術（portacaval shunt）一度是治療肝硬化的常用方案。一九六六年有人開始進行研究時，發現一個有趣的現象。在沒有控制組的研究中，有96.9%的研究發現，醫師斷定這種治療具有中等以上程度的有效性；在有控制組但沒有隨機分派的研究中（並不符合眞實驗設計的要求），有86.7%得出同樣的結論；而在有控制組又有隨機分派的研究中，只有25%得到同樣的結論。可見，這項治療的效果——現在我們已經知道無效——被缺乏控制的研究資料極誇張地高估了。羅斯和尼斯貝特也指出：「非經標準實驗程序所獲得的結果，是一種『安慰劑效應』或是因沒有隨機分派而產生的偏差。」(p. 207) 他們還討論了在無隨機分派的情況下，選擇偏差（見第五章的討論）如何產生虛假相關這一問題。比如說，如果選擇治療的那些病患具有「典型樣本」的特點，或其家屬態度積極、更關心他們的病情，那麼，就算結果發現治療組與控制組有差異，也不能說明是治療本身有療效。

作結論前有必要先獲取一些比較資訊，這種思維習慣似乎不是與生俱來的，這也就是爲什麼所有科學訓練中都包含強調設計控制組重要性的研究方法課程。然而，控制組的具「非鮮明性」——除了改變某個關鍵因素外，控制組與實驗組完全接受同樣的待遇——使得我們從表面上很難看出它有多麼重要。人們爲什麼不重視比較（控制組）的資訊，心理學家在這方面做過不少研究。例如，

在很多研究中（Shanks, 1995; Stanovich & West, 1998），我們給受試者呈現一個2×2的實驗資料矩陣：

	有療效	沒有療效
治療組	200	75
無治療組	50	15

表中的數字代表每種情況的人數。具體地說，這個表告訴我們有二百人在獲得治療的條件組中接受治療後呈現病情好轉的現象；七十五人沒有好轉；五十人是在沒有接受治療的條件組，後來卻也呈現病情好轉的現象；另有十五人既沒接受治療也沒好轉。研究者讓看過這一矩陣的受試者判斷這個治療的有效程度。結果發現非常多人認爲有效，甚至有相當多人認爲非常有效。因爲他們首先注意到的是人數最多（二百人）那一組，正是接受治療並見好轉的。接著，他們注意到獲得治療的治療組，治後呈現好轉的人數（二百）比沒有呈現好轉的（七十五）多得多。所以就下了結論，認爲這一治療是有效的。但是實際上，這一治療方案根本就沒有效果。

此話怎講？要弄清楚爲什麼這個治療無效，我們必須重點地看一下代表沒有接受治療的控制組（沒有獲得治療的人）的兩格資料。我們可以看到，控制組的六十五人中有五十人，或76.9％，即使沒有得到治療卻也呈現病情好轉。與二七五人中有二百人，或72.2％，在接受治療後呈現好轉相比，很顯然，控制組的好轉率還大於治療組！這就表示，這種治療總的來說是完全沒有療效的。可見，由於經常忽視無治療控制組的結果，並且只將注意力放在治療組資料中的最大值和好轉率上，使得許多人認爲這一治療是有效的。總而言之，人們很容易忘記：控制組

的結果，是在我們對治療組結果進行解釋時，非常關鍵的一個背景資訊（參見 Doherty, Chadwick, Garavan, Barr, & Mynatt, 1996）。

◎神馬聰明漢斯的個案

行爲科學中有個著名的故事，很可以說明利用實驗控制消除對現象的各種不同解釋的必要性：聰明漢斯的故事，牠是匹會算數的馬。八十多年前，德國一名中學教師向大家介紹了這匹會算術的馬。當漢斯的訓練師向牠呈現加、減或乘法運算的題目時，牠會用蹄子敲出令人驚奇的準確答案。

許多人被聰明漢斯的表現驚呆了，而且迷惑不解：難道牠眞的能做過去人們認爲馬不能做的事嗎？你可以想像當時人們的驚訝。對漢斯獨特能力的強力見證受到德國媒體的廣泛報導，柏林的一家報社記者說：「這匹會思考的馬將使科學界在很長的一段時間做許多的思考。」（Fernald, 1984, p. 30）此一預言很快成爲事實，儘管結果不是那位記者當初預期的。一組「專家」對漢斯進行觀察並證明牠確有此一能力。因此所有人更困惑了。只要不實施控制比較研究，僅通過對孤立事件的觀察，困惑就將一直持續下去。所幸心理學家方格斯特（Oskar Pfungst）對漢斯的能力重新進行系統研究，謎團才得以解開（Spitz, 1997）。

方格斯特繼承實驗設計的優良傳統，系統地操弄動物表演的環境。這樣就可以創造出一些「人工」環境（詳見第七章），從而以之驗證可以解釋漢斯之表現的各種說法。經過許多仔細的測試，方

格斯特發現這匹馬確實具有一種特殊能力，但並不是算數能力。事實上，這匹馬更接近一位行爲科學家，而不是數學家。這是怎麼回事？原來，漢斯是非常用心審視人類行爲的觀察者。當牠要敲出數學答案時，會觀察訓練員或其他提問者的頭。當漢斯快接近答案時，訓練員會不自覺地歪下頭，這時漢斯就會停下來。方格斯特發現馬對視覺所提供的線索非常敏感，牠可以察覺非常細微的頭部運動，於是方格斯特想出另一個測試馬之能力的方法：讓不知道問題答案的人呈現問題，或讓訓練員在馬的視野以外呈現問題，結果，漢斯便失去了牠的「算術能力」。

漢斯的故事是個很好的例子，它向我們揭示愼重區分對現象的**描述**和對現象的**解釋**之重要性。馬能夠敲出訓練員提出的算數題之答案是不爭的事實，訓練員也沒有撒謊，許多旁觀者也見證了這匹馬確實準確地敲出訓練員所提出算題的答案。問題出在下一步：由馬能敲出正確答案推論出牠具有算術能力！的確，下結論說馬具有算術能力是對現象的一個**假設解釋**，而且是衆多可能解釋這一現象的說法之一。但是從該馬能正確地敲出算數題答案就很快地推論說牠具有算術能力，是言之過早。這一解釋必須先付諸實徵驗證。一旦付諸實徵驗證後，就證實是完全沒有根據的。

在方格斯特進行研究之前，對漢斯進行觀察的專家犯了一個根本的錯誤：他們沒有察覺馬的神奇表現可以有許多不同的解釋。他們認爲，只要訓練員沒有撒謊，漢斯又的確敲出算數題的正確答案，就足以說明馬具有算術能力。方格斯特對此卻進行更科學的思考，他意識到這一解釋僅是對該馬的行爲的衆多可能解釋之一，有必要設計受控制的實驗情境以區分各種不同的解釋：該馬有算術

能力，或是該馬有覺察力。於是方格斯特設計一個讓訓練員在隔板後面向馬提問的情境，以區分兩種可能的解釋：該馬具有算術能力，或者該馬是依據視覺線索做出反應。如果漢斯確實具有神奇的算數能力，那麼讓訓練員在隔板後面提問，對牠的表現應該不會產生任何影響；反之，如果該馬只是依賴視覺線索做反應，那麼訓練員在隔板後提問就會影響牠的表現了。一旦後者出現，方格斯特就可以排除該馬具有算數能力的假設（Spitz, 1997）。

這裡可以和第三章講到的節約原則聯繫起來，這一原則指出，當兩個理論具有相同的解釋能力時，較簡單的（涉及較少的概念及概念之間的相關）那個理論是比較可取的。這裡存在爭論的兩個理論——一個認爲漢斯眞的具有算數能力，另一個認爲牠在辨別行爲線索——在節約原則上是截然不同的。後者不需要對先前的心理學及大腦理論做出巨大的調整，僅僅要求我們對馬在視覺線索上可能存在的敏感性（現在看來這一解釋的可能性很高）之看法稍加調整即可。而前一理論——認爲馬眞的能運算——則需要我們對演化論科學、認知科學和腦科學等方面的十來個概念進行較大的修改。這一理論顯然是極端的不節約經濟。因爲，如果我們認爲這個理論正確，許多其他科學的概念就要跟著徹底的改頭換面（這就是第八章將討論的關聯原則）。

◎一九九〇年代的聰明漢斯：一個悲劇的衍生

聰明漢斯的故事經常在方法課上用來說明必須採取實驗控制這一原則的重要性。不幸的是，人

們不僅有時忽視了聰明漢斯個案的一般重要性，即使是其直接教訓看來也被忘得一乾二淨了。在整個一九九〇年代早期，全世界的研究者都在驚恐中觀望——就像觀看汽車撞毀的慢動作——眼看著聰明漢斯的一個現代版在眼前發生，並且導致悲劇性的結局（Gardner, 2001; Spitz, 1997; Twachtman-Cullen, 1997）。

自閉症（autism）是一種嚴重的成長時期障礙，典型特徵有：人際交往有缺陷、嚴重語言發展遲滯及異常、活動及興趣的範圍狹窄等等（Frith, 1989, 1993）。由於許多極度缺乏人際溝通能力的自閉症兒童外表都很正常，致使父母特別不能接受他們的子女有病。這就不難想像，為什麼在一九八〇年代末九〇年代初，當自閉症兒童的父母聽說澳洲有人發明一種可以使過去從不用語言溝通的自閉症兒童與人進行語言交談的技術時，為之興奮不已了。這種技術叫做**溝通輔助**（facilitated communication）。當時，一些極具影響力的媒體及節目，像《六十分鐘》（*60 Minutes*）、《花車遊行》雜誌（*Parade*）及《華盛頓郵報》（*Washington Post*）等，都對此進行毫無批評性的宣傳報導（參見 Dillon, 1993; Jacobson, Mulick, & Schwartz, 1995, 1996; Mulick, Jacobson, & Kobe, 1993; Spitz, 1997; Twachtman-Cullen, 1997）。這個技術的發明者宣稱，自閉症患者以及其他因成長性障礙導致不能講話的兒童，只要把手和胳膊放在這部具有同情心的「輔助器」上，就可以在它的幫助下，在鍵盤上敲出相當有文采的句子來。可以想像，當部分先前語言能力極差的自閉症兒童做出這樣驚人的語言表現時，他們深受挫折的父母會受到多大的鼓舞。發明家還聲稱，這個技術還可以幫助有嚴重智力障礙的失語兒童。

做父母的這種興奮很容易理解，但是專業人員也這麼輕易接受這個發明，卻非常令人費解。不幸的是，在尚未展開對這一技術的療效進行任何控制研究之前，其功效已經透過大量的媒體傳播給許多滿懷希望的父母。如果有關的專業人員在實驗控制上稍稍具有基本訓練，就應該馬上意識到它與聰明漢斯個案之間的相似性。這種輔助器可以說是一個永遠關注兒童成功的、富於同情心的人，在輔助過程中有許多機會，在意識或潛意識的層次，引導兒童敲擊鍵盤。其實，另一個觀察到的現象，也支持被輔助器所給的線索引導這一解釋：兒童有時在不看鍵盤的情況下，就可以打出複雜難懂的句子。另外一個支持這個解釋的觀察是，一些從沒有見過字母的幼童，也可以用這個輔助器創作出具有深度技巧的英文散文。

一些有控制的研究報告聲稱，他們透過適當的控制對受輔助的溝通技術進行驗證，而每個研究都清楚地得到同一結論：自閉症兒童的表現，依賴於輔助器發出的、不易被察覺的線索（Beck & Pirovano, 1996; Burgess, Kirsch, Shane, Niederauer, Graham, & Bacon, 1998; Cummins & Prior, 1992; Hudson, Melita, & Arnold, 1993; Jacobson et al., 1995, 1996; Levine, Shane, & Wharton, 1994; Prior & Cummins, 1992; Shane, 1993; Simpson & Myles, 1995; Smith & Belcher, 1993; Spitz, 1997; Wheeler, Jacobson, Paglieri, & Schwartz, 1993）。這些研究中，有些運用的控制都與聰明漢斯這個典型例子中的相似。他們設計的控制情境是給兒童和輔助器各自呈現一個畫出來的東西，但彼此看不到呈現給對方的東西是什麼。當向兒童和輔助器呈現同一個東西時，兒童打出的是自己看的東西之名字；但是當給兒童和輔助器呈現不同的東西時，兒童

打出的卻是呈現給輔助器看的東西之名字，而不是自己正在看的。也就是說，兒童的回答取決於輔助器而不是自己的能力。得知溝通輔助技術不過是個聰明漢斯現象，而不是突破性治療技術的結果，這個結論並沒有使研究者爲之雀躍。因爲這一聰明漢斯現象不只在輔助器之療效研究中發生，許多其他研究都有同樣的問題。心理學家康明斯和普萊爾（Robert Cummins & Margot Prior, 1992）下結論說：「很明顯的，許多研究助理都是運用不易察覺的視覺線索或自身的動作姿態等，來控制他們病患的反應，從而得到有療效的資料。這是目前研究證據所給我們的不想要的、卻又不可避免的結論。」（p. 240）

這一不幸的事件愈演愈烈。在一些治療中心，有當事人在接受輔助器幫助的溝通過程中，講出過去遭到父親或母親性虐待的經驗（Dillon, 1993; Seligmann & Chideya, 1992; Spitz, 1997; Twachtman-Cullen, 1997）。這些孩子被迫從家中搬出來，直到這種指控證明是無中生有之後，才再搬回家：「許多法院的訴訟案件，皆因溝通輔助技術引起的，其中最引人注目的是幾起與虐待有關的案件。當法院下令進行的查證證實與輔助器嚴重影響人們的溝通有關之後，這些指控就都取消了。」（Smith & Belcher, 1993, p. 176）。赫德森等人（1993）報告了對一名婦女的測驗。這位二十九歲婦女被認定智力極度低下，一直和父母生活在一起。在接受溝通輔助治療期間，她報告曾遭到生活中重要人物性侵犯。在進入法律程序時，赫德森等人運用先前研究者使用聰明漢斯研究中所用的邏輯，重新測試該婦女在接受輔助器治療時的溝通能力。透過兩個不同的耳機，向她和她的輔助器讀出問題。當兩者

被問到的問題相同時，每次她都能夠正確的回答；但是當問題不同時，她從沒有給過正確的答案，而且其中40％回答了輔助器被問的問題。

就像《新聞週刊》報導的：

這種未經證實的控訴已經造成很大的傷害。父母不得不在忍受鄰居的責難和聘請律師的高額費用下，讓人把自己的孩子帶走，安置在臨時監護人的家中——而最後換來的不過是學校幾個月後寄來的一封道歉信。受到牽連的老師及其助手，也因為這一醜聞的陰影而斷送事業。（Seligmann & Chideya, 1992, p. 75）

當愈來愈多的證據發現這種缺乏實徵基礎的治療不只是無害而已時，人們終於在媒體的喧擾之上，開始注意到專家的意見。（「呃，也許它是有效的，但是假如它是無效的，那怎麼辦？」）推廣這種尚未證實的治療技術所帶來的社會成本，卻是千眞萬確的。波士頓兒童醫院（Children's Hospital）的溝通促進中心（Communication Enhancement Center）主任夏恩（Howard Shane, 1993）平靜地說：

所有的科學跡象都表明，溝通輔助技術是無效的。……由於這種方法可能帶來的潛在危害，所以我們很難建議繼續使用它。例如，在接受輔助器治療期間發生的性虐待指控，不當的教學人力資源安排，以及為了啟用這項無效技術而撥出的鉅額培訓和研究經費。

（p. 11）

俄亥俄州立大學（Ohio State University）兒科及心理學教授繆利克（James Mulick）（參見 Mulick, Jacobson, & Kobe, 1993）更一針見血地指出這種一窩蜂所付出的代價：

> 如果沒有對FC（溝通輔助技術）的大力宣傳，我們可能可以得到更多的人力和金錢用於發展基於更有實徵基礎的、更可行的長遠策略，以解決困擾兒童的這一問題。FC的擁護者為研究和專業文獻所帶來的理論上的混亂，損害了對能力缺陷及其成因方面的知識之累積，也引開那些真正有心有力想將成長障礙的研究加以綜合的人之注意力。FC引起的混亂，甚至使其他已成功運用於殘障人士的非言語溝通系統也普遍遭到大眾的排斥……經驗告訴我們，殘障人士可以成為家中有大用的成員，而無需求助於神醫神藥。他們可以尋求現有的、有科學依據的有效幫助。具有科學訓練、富有熱情的專業人士的努力，絕對、也永遠比各種瞬間即失的一窩蜂式治療好得多。治療的進步及知識的進展，是建築在嚴格的訓練、對正確及科學標準的尊重，及對治療療效的客觀論證上。（pp. 278-279）

以上這個例子再次說明了，因迷信見證敘述可能造成的傷害，以及認爲一窩蜂式治療與偽科學是無害的這個觀點所可能帶來的謬誤（詳見第四章）。由此我們也可以看到，想正確地解釋行爲，實

驗方法中的控制和操弄是必要的、也是不可替代的。

這裡需要再次強調節約原則與控制和操弄的相關。如果患有嚴重語言障礙的自閉症兒童可能透過一顆百發百中的「魔術子彈」（見第九章）之千預就能治癒，那麼它顯然可以推翻過去幾十年來對患自閉症兒童所做的認知、神經心理學以及大腦特徵等領域的所有研究成果（Baron-Cohen, 1995; Frith, 1993; Leslie, 2000; Sigman & Ruskin, 1999; Yirmiya, Erel, Shaked, & Solomonica-Levi, 1998），從而我們必須大大修正已知的認知和生物知識。溝通輔助技術與其他科學研究成果幾乎完全沒有相關性（見第八章）。

◎分離變數：精心設計的情境

古德伯格的例子給我們上了重要的一課，有助於消除科學進程中的許多錯誤觀念，特別是在心理學的研究方面。世界上發生的許多事經常與很多因素相關，為了仔細區辨許多同時發生的事件之因果影響，我們必須設計出現實世界中不曾有的情境。科學實驗就是要打破自然界原有的各種相關性，將單一變數可能造成的影響分離出來。

心理學家採用的也是同樣的方法：透過操弄和控制以分離變數。例如，對閱讀過程感興趣的認知心理學家，已經在研究是什麼因素使人對文字的感知更容易或更困難。毫不奇怪地，他們發現長的單詞比短的更難確認。乍看之下，我們可能認為字的長度應該很容易測量：只要蒐集兩組單字，一組長一組短，然後測量受試者在閱讀這兩組字時的速度差異就可以了。但說來容易做來難，長字

在語言中出現的頻率小，而頻率本身就會影響我們對字的感知。因此，長短詞中出現的任何不同，既可以歸因於長度，也可以歸因於單字的出現頻率，或是兩者的聯合效應。爲了檢視在不受頻率影響的情況下字長是否影響感知，研究者就必須構建特殊的單字組，以確保字長和頻率不會同時發生變化，而同時影響對字的感知。

與此相似，古德伯格之所以能對原因做出有力的推理，是因爲他設計了一系列在自然環境中不會發生的特殊情境（回想一下，他的一個實驗操弄是要受試者吃下人體排泄物，這是何等的不自然呀！）。回想方格斯特同樣地爲測試聰明漢斯的能力而設計的一些特殊情境，包括提問者不知道問題答案的情境。許多人只觀察到馬在正常情境下回答問題（亦即在提問者知道答案的情況下），但從來沒有想過馬怎麼會做出這樣驚人的表現，反而做出馬眞的具有算數能力的錯誤判斷。

同樣的，在檢驗溝通輔助技術之功效時，研究者也必須設計一些特殊的情境。展示給輔助器和給兒童的刺激物必須分開呈現，並且不讓任何一方知道對方的刺激物是什麼。這種特殊情境，對檢驗解釋同一現象的不同假設至關重要。

許多心理學的經典實驗都包含這種分離現實中之自然聯結的邏輯，以判斷哪一變數是現象的決定因素。心理學家哈洛（Harry Harlow）的著名實驗（Anderson & Anderson, 1996; Harlow, 1958）提供了這樣一個相關的事例。當時，關於母嬰依附現象有一種流行的假設：這種依附是因爲母親是嬰兒食物的來源。然而問題是，母親能提供的當然不只是食物（還有舒適、溫暖、愛撫、刺激等），也有可能是

這些其他的原因造成此一依附。哈洛（1958）想對這一問題進行驗證，看看到底哪個才是造成依附的眞正原因。他觀察了猴子在不同研究情境裡的行爲。在這些情境裡，除了一個可能造成母嬰依附的原因變量在變化之外，其他變數都受到了控制，所有情境裡都保持不變。在那一原因變量上則讓猴子選擇與用不同材料做成的「人造」母猴中一起生活。哈洛發現，絨線做成的人造「母猴」給予的觸摸舒適遠勝於鐵絲網做成的「母猴」。在嬰猴兩個星期大時，牠更願意接觸冰冷的絨線母猴而不是溫暖的鐵絲網母猴，顯示觸摸所帶來的舒適感遠比溫暖感更重要（Harlow & Suomi, 1970）。最後，哈洛還發現，即使當嬰猴的食物完全由鐵絲網母猴提供，牠仍選擇絨線母猴。因此可以證明，依附僅歸因於母親能提供食物的假說是錯誤的。哈洛能做出這些結論，是因爲他可以把現實世界中同時變化的變數分離開。

總之，對科學家來說，檢驗解釋某一現象的特殊假設或理論時，有必要設計一些特殊的情境，僅僅在現實世界的自然狀態下觀察事物是遠遠不夠的。人們觀察下降和移動的物體已經有幾個世紀了，卻沒有人得出移動和重力的正確規則和定律。直到伽利略和其他科學家設計了一些人爲情境觀察物體運動時，才得到眞正能解釋運動的定律。在伽利略的時代，很少有人會觀察光滑的銅球從光滑的斜板上滑下。儘管在這個世界裡有許多運動發生，但很少是以這種特殊形式出現的。然而，正是這樣的一個非自然情境，加上其他類似的情境，促使我們發現關於運動和重力的第一個眞正解釋性的定律。講到運動定律，你不是在本章開始做了個小小的測驗嗎？

◎直覺式物理學

其實，本章開頭提到的三個問題來自約翰霍普金斯大學（Johns Hopkins University）的心理學家麥克勞斯基（Michael McCloskey）的一本著作。麥克勞斯基（1983）研究過他所稱的「直覺式物理學」，亦即一般老百姓對於物體運動的認識。有趣的是，這些認識經常與物體的眞實運動完全相反（Catrambone, Jones, Jonides, & Seifert, 1995; diSessa, 1996）。

比如說，在第一個問題中，一旦繫住正在轉圈的球之繩子被切斷，球會沿與繩子呈90度（圖的切線）角的直線飛出去。麥克勞斯基發現，三分之一的大學生錯誤地認爲球會沿拋物線飛出去（同時參見Catrambone et al., 1995）。對於轟炸機飛行員的問題，大約一半的人認爲炸彈應該在目標的垂直上空拋出，這足以看出人們對物體初始運動決定其後來運動軌跡這一知識有多缺乏。實際上，炸彈應該在飛機距離目標還有五哩時拋出。受試者的錯誤，原因並不是這些題目都是假想的抽象情況所以很難回答。因爲，當要求他們在屋子裡邊走邊丟一個高爾夫球時，幾乎一半以上的人根本不知道球在下落時會向前移動。至於最後一道題的答案，許多人都沒有想到，子彈從槍管中發射出來，與子彈從同一高度下落會同時到達地面。

你現在可以算一下自己回答這幾個問題的成績了。如果你最近沒有上物理課，有可能至少錯一題。「物理課！」你可能會抗議：「我最近當然沒有學過物理學，當然答不對了！這個測驗太不公

平呀！」但是，喂，你爲什麼需要上物理課才能知道這些題目的答案？在你的生活中，看見物體從高處掉下恐怕不下幾百次了吧？而且它們都是在**自然發生**的情境下掉落的。每天在生活中，你也見過不少物體在運動，而且你是在「眞實」狀態下看到的。所以，你當然不能說從未看過正在運動或正在下落的物體。就算你辯稱從未看過像子彈飛出一樣的事件，但是我們幾乎都看過小孩將旋轉的物體放開的情境，而且很多人也見過從飛機上扔東西的情境。你說你從未見過完全類似的情境，似乎有點強詞奪理吧？既然你每天都會遇到這些運動或下落的物體，爲什麼還是不能準確的預測那些與常情稍有不同的情境會發生什麼事情呢？

在理解科學家爲什麼要那麼做這方面，麥克勞斯基的成果向我們展示了一些根本要素。不管人們對運動和下落物體經歷再怎麼多，對運動的直覺理論仍非常不準確。關鍵是我們要理解外行人之所以會有那些錯誤的觀念，正是因爲他的觀察是在「自然的」、而不是用科學家的方式在控制的情境中進行的。因此，如果你在本章開頭的小測驗中答錯了，也不要感覺自己無知或不夠好。要知道，一些世界上偉大的思想家觀察物體下落好幾個世紀了，但是對運動物理都沒能想出什麼比高二學生更準確的理論。在《科學人》（*Scientific American*）雜誌上的一篇文章中，麥克勞斯基（1983）觀察到，他的很多受試者對運動的物體所持的不正確看法，竟與牛頓時代之前三個世紀的想法不謀而合。麥克勞斯基的現代受試者和中古世紀的哲學家在某些方面是相同的：兩組人都是在自然世界中對物體運動做觀察，而沒有運用人爲設計的情境進行科學的操弄、控制和比較。

即便是大量的個人經歷，也不能防止對物體運動本質是什麼的錯誤觀念的產生。飛行員蘭格威士奇（William Langewiesche, 1993）在闡述關於飛機轉彎角度這一知識的發展史時指出，二十世紀早期的飛行員，對使用旋轉器這類裝置非常排斥，因爲他們相信「本能的平衡」。然而，在多雲的天氣裡，因能見度低，飛行員對陸地及岸邊的感知不足時，這些「本能」會因爲不能借助飛行員的這一直覺，而無法正確地將飛機「平衡」到想要的角度。許多飛機墜毀及險些墜毀的事件之後，飛行員終於領悟到一個沈痛的教訓：再多的直覺，也不能取代飛行物理學的知識（Langewiesche, 1993）。

◎直覺式心理學

哲學家丘奇蘭德（1988）辯稱，如果我們關於運動物體的世俗理解都不正確，那麼很難相信我們對有關人類行爲這一更複雜的領域之世俗見解會是正確的：

我們最初關於運動的世俗理解是非常模糊及混亂的，而且最終將被更成熟的理論完全取代。我們早期關於天堂的結構及活動的世俗理解也是完全離譜的，它們之所以依然存留下來，只不過是作為一些歷史的教訓，讓我們知道自己可以錯到什麼地步。我們關於火的本質以及生活本質的世俗理解也同樣愚蠢。既然過去許多世俗想法都被推翻了，所以我們可以這樣一直數落下去……，但是有關人類的心智活動這一現象，要比前述任何現象複雜

> 和難以理解得多。如果我們對其他事情的理解都這麼糟糕，那麼，想在一開始就正確地認識心理學知識，簡直就是天方夜譚。（P. 46）

> 生物學家威爾遜（E. O. Wilson, 1998）在贊同丘奇蘭德的推測的同時，指出「大腦這部機器，不是爲了理解它自己，而是爲了延續生命。這兩個目的根本上完全不同，如果沒有從科學中獲得的事實知識的幫助，大腦也只能看到世界的一些小碎片。爲了考慮明天如何繼續生存，它必須把注意力放在必須知道的世界的一角……這就是爲什麼即使到今天，人們了解他們的汽車比了解他們的大腦還多——以及爲什麼對於人類思維的基本理解還是停留在一種實徵層次，而並沒有進升到哲學或宗教的層次。」（pp. 96-97）

審視有關人們行爲理論的現代文獻，我們發現丘奇蘭德的推測是正確的。在第一章，我們舉例說明了一些對人類行爲的普通常識性（或世俗性）理解是錯誤的。在科恩（Alfie Kohn）的著作《你知道他們說些什麼……》（*You Know What They Say...: The Truth about Popular Beliefs*, 1990）和薩拉（Sergio Della Sala）的著作《心智的迷思》（*Mind Myths: Exploring Popular Assumptions about the Mind and Brain*, 1999，中譯本心理出版）中有更多這方面的例子。儘管兩本書都是爲一般大眾寫的，他們還是大量採用有同儕審核的科學學術雜誌上發表的論文，做爲結論的證據。兩本書中討論的許多通俗理論都與人類行爲有關，例如，沒有證據顯示宗教信仰虔誠度高的人比虔誠度低的更有利他行爲的傾向（Paloutzian,

1983; Smith, Wheeler, & Diener, 1975）。許多研究顯示：宗教信仰的虔誠度與其參與慈善活動、幫助處於困境中的人或不欺騙他人等行爲傾向之間，沒有簡單的關係（Paloutzian, 1983）。確實，在大量研究文獻中，沒有跡象顯示高宗教虔誠度的人比自認爲無神論者更有善心或更願意幫助別人。

有關民間信念不正確的例子可以舉出一大堆：例如，許多人認爲月亮的盈虧影響人的行爲，其實這是不對的（參見 Byrnes & Kelly, 1992; Coates, Jehle, & Cottington, 1989; Culver, Rotton, & Kelly, 1988; Rotton & Kelly, 1985）。有些人認爲「異性相吸」，他們也錯了（參見 Buss, 1985; Buss & Barnes, 1986; Murstein, 1980）。有些人認爲「熟悉產生鄙視」，這也是錯的（參見 Bornstein, 1989）。還有人認爲盲人天生有超常敏感的聽覺，情況並不是如此（參見 Niemeyer & Starlinger, 1981; Stankov & Spilsbury,1978）。這類的例子舉不勝舉。有興趣的讀者可參考科恩（1990）的書，他在書中詳細列舉了大量沒有實徵證據支持的有關人類行爲的民間信念（當然，這裡列舉的所有結論都是建築在機率的基礎上，詳見第十章）。

人們有關行爲的直覺理論的許多不足之處，說明了爲什麼我們需要有控制的心理學實驗：這樣我們才能走出對人類行爲的古老想法，走向建立更準確的科學概念。

摘要

實驗方法的中心思想是操弄及控制。在相關研究中，研究者僅僅觀察兩個自然變數間的變化是

否存在關聯；而在眞正的實驗中，研究者操弄那個假設可能是一個現象的原因之變數，並透過控制和隨機分派維持所有其他可能與該原因變數相關的其他變數恆定，以尋找原因變數對現象變數的眞正影響。這種方法排除了相關研究中出現的第三變數問題。第三變數問題的出現是因爲在自然世界中，許多不同事物是相互關聯的。實驗法可以視爲用來分離這些自然相關的一種方式。之所以這樣做，是因爲透過操弄假定的原因變數和恆定所有其他變數，可以將假定的原因變數分離出來，單獨審視它對現象變數的影響。但爲了分離這些自然相關，科學家經常必須設計一些在自然世界中找不到的特殊情境。

第七章 「但這不是眞實生活啊！」

——「人爲性」的批評與心理學

上兩章中，我們闡述了心理實驗的基本邏輯思路。現在來看看我們經常聽到的，針對心理學的種種批評。透過對這些批評的討論，我們可以加深對科學方法的理解，而本章特別要深入探討的是，因爲心理實驗是人爲的或不是「眞實生活」，而否認實驗法有用之類的批評。理解這類批評不能成立的原因，更有助於我們了解心理實驗，進而了解心理學。

自然性有時不是必需的

經過第六章的討論，我們已經很清楚這一點了吧。正如前面提到的，科學實驗的人爲性不但不是什麼毛病，恰恰相反，正是這人爲性才使科學方法擁有解釋自然世界的巨大力量。與一般的認知

不同，科學實驗的人爲性是科學家深思熟慮後的結果，而非一時疏漏。科學家**精心**設計一個非自然的環境，以便將眾多可能決定事件發生的相關變數分離出來。

有時，起決定作用的條件在自然狀態中已經存在，如史諾與霍亂的例子，但這種情況畢竟只是少數。更多時候需要科學家透過一些自己設計的，有時甚至是新奇的方法，才能操控事件的發生，如古德伯格與癩皮病的例子。在許多情況下，這些操控手段不會在自然環境下發生，因此科學家必須將之引入實驗室中，進行更精密的控制。例如，在重力與運動的早期研究中，爲了製造某種特殊情境以便觀察物體運動，就構建了許多很特別的物體。這就說明，爲了分析某種現象，在實驗研究中設計一個**非自然**且極端的情境是很有必要的。

實際上，如果科學家嚴格信守「自然」觀察，那麼許多現象就完全沒有發現的可能。例如物理學家爲了研究物質的根本特性，造出長達一哩、巨大無比的加速器，用來誘發微粒子之間的碰撞，在這些碰撞過程中，他們發現一些存在時間不到十億分之一秒的未知粒子。這些新粒子的特性對原子結構的理論有重大意義。但在自然情境下往往不存在；即使存在，也不可能用自然觀察法察覺得到。很少有人對物理學家的這種做法提出懷疑，用特殊的、有時甚至稀奇古怪的方法探究自然，被視爲深入了解宇宙的有效途徑。然而，一旦應用到心理學，就被看成是無效的。

如果人們不了解在實驗室中設計特殊情境的重要性，往往會誤解科學家的做法，而心理學家更是處在誤解的重災區。當心理學家拿出某一行爲在實驗室中所蒐集到的證據時，經常會聽到外行人

這樣的悲歎：「但這不是眞實生活啊！」深入探討這句話，會發現這些人堅持只有在自然情境下做的研究，才是眞正的研究，而心理學的實驗研究是怪異的，因此也意味著心理學不足以作爲一門科學。

人們一般不會意識到，心理學家採用的看似奇怪的手段，並不單只心理學才用，其實這些手段只是將科學方法應用在研究人類行爲方面而已（Banaji & Crowder, 1989）。這些爲了了解世界而使用的奇特方法，同樣會在任何一門科學中看到。不幸的是，心理學家卻因此受到雙面夾擊！除了控訴心理學家不用自然情境研究行爲，因此不能反映眞實生活，還有一些同樣因爲對研究方法的無知，而認爲心理學永遠不能成爲科學的人，因而造成對心理學家的藐視，雖然心理學家也像其他科學的研究者一樣，透過設計必要的、特殊的情境以發掘對某一現象更有力、更精確的解釋，他們卻視而不見。

對科學研究要求只局限於觀察眞實生活，會妨礙我們發現許多事情。例如，現已廣泛應用於衆多領域，如週期性偏頭痛、緊張性頭痛的控制、高血壓的治療及放鬆訓練等的生物回饋（biofeedback）技術（Miller, 1985a），一開始源於一個實驗研究結果：當人們能夠透過一個監控器得到自己內部生理機能運作的回饋，就能學會控制這些過程的一部分。當然，由於人類本身沒有這種透過外部回饋監控內部生理機能的器官，所以，這種控制能力只有在特殊的實驗情境下才會表現出來。如果只通過自然觀察研究這個問題，就永遠不會發現此一現象了。正如安德森、林賽和布希曼（C. A.

Anderson, J. J. Lindsay, & B. J. Bushman, 1999）所說的：「大多數實驗研究的目的，都是爲了發現一些概念變數之間有哪些理論上的關係，而這些變數又是在眞實世界中不能充分分離出來以便作精確審察的。」（p. 4）

◎對「隨機樣本」的誤解

有時，人們對「但這不是眞實生活啊！」的抱怨，是源於對心理實驗目的的誤解，其實它並不難理解。媒體的宣傳使大家對調查研究都很熟悉，尤其是正式選舉中的民意調查。並且也漸漸意識到選舉民調中的一些重要特性，特別是媒體給予格外關注的隨機取樣與樣本代表性問題。人們知道這兩個特性是民意測驗準確度的關鍵，也因此使許多人誤以爲隨機取樣及代表性是所有心理學研究的必要條件。由於心理學研究中很少使用隨機取樣選擇受試者，那些認爲只有使用隨機取樣才行的外行人就認爲，心理學研究由於沒有這麼做，因而沒有反映眞實生活，也因而是沒有用的。

但只要想一下其他科學的情況，就很容易理解這種想法的荒謬。化學家從未試過將化合物隨機取樣；生物學家不會做什麼關於細胞與器官的隨機取樣；在醫學研究中應用的老鼠和猴子，也不見得就能代表牠們的物種：這些研究都是在與自然環境完全不同的實驗室中進行的。實際上，這些研究情境可以說絕對是獨特的，然而這些研究的結果，卻可以幫助我們了解生物學。多數心理學的研究也一樣，並不是每個研究都必須用隨機取樣找研究對象。

◎理論導向研究與直接應用的異同

那爲什麼對選舉來說隨機取樣這麼重要，對心理學研究而言就不那麼重要呢？這個問題，我們可以透過考察各種應用研究與大部分基礎研究之間的不同以尋找答案。維吉尼亞大學的心理學家默克（Douglas Mook, 1983, 1989, 2001），根據不同研究類型所要求的不同預測層次對上述問題進行闡述。許多類型的應用研究，目的就是要將研究結果直接應用到日常生活中的一個特定情境。對應用研究而言，預測必須是：研究與現實生活有一對一——即默克稱作「類比」——的關係。也就是說，研究結果必須能直接應用才行。選舉中的民意測驗便是一個應用研究做「類比」預測的例子，目標就是爲了預測選舉當天的投票結果。由於研究結果要直接應用，樣本的隨機取樣性及情境的代表性就顯得格外重要。

儘管應用心理學研究已經愈來愈普遍，但如果由此認爲這類研究是典型的心理學研究則是錯誤的。實際上，大多數心理學研究（其他學科其實也是如此）目的並不在於應用，它關心的預測層次都是從理論到特定的研究情境，而且得到的結果也不能直接應用，必須透過對理論的調整，再將理論和其他科學規律結合，才能應用到實際問題中。簡而言之，大多數理論導向的心理學研究，是爲了驗證有關某一心理歷程的理論，而不是把研究結果推廣到現實中的某個特殊情境中。

目的主要在於驗證理論的研究，通常稱爲**基礎研究**。應用研究的目的是從研究資料直接走到現

「但這不是真實生活啊！」

實世界中，基礎研究則關注理論的驗證。但是，如果僅僅根據一個研究是否能實際應用來分辨這兩種研究的差別，就可能太簡化了（參見 Nickerson, 1999，有關兩者的相互依賴性）。因爲這一差別常常是時間的差別造成應用有早晚的問題。應用研究的發現可以立即利用，然而沒有任何東西比一個普遍的、精確的理論更有用，只不過這種理論需要時間的醞釀。在科學歷史的長河中，有很多的例子讓我們看到，儘管原本一個科學家並無意解決什麼實際問題，但他的理論或發現，最終還是解決了真實世界中的很多問題。

在布魯克哈芬國家實驗室（Brookhaven National Laboratory）這個物理研究龍頭機構中工作的克瑞斯和薩米爾斯（Robert Crease and Nicholas Samios, 1991），曾在其文章中講述幾個基礎研究本身具有應用價值的例子。他們提到倫琴（Wilhelm Roentgen）的故事。倫琴「爲了弄明白爲什麼放在他儀器旁邊的一個螢幕會意外地自動發光，結果發現自然界的一個新現象——他稱之爲X光。三個月之後，X光就被用來檢查骨折了。」（p. 82）再看看弗羅瑞（Howard Florey）的例子，他和一名同事從事抗菌機制的調查，「在他們要研究的微生物清單中有青黴素，這種細菌在幾年前才偶然發現，但它的抗菌能力卻一直沒有人注意到。由此可見，現代醫學的進展很多都是建立在這類對應用沒有特別興趣的學術研究的發現和發展之上的。」（p. 82）

凱特（Seymour Kety, 1974）也談到他如何在看似毫無相關的科研中發現治療精神分裂症的特效藥氯丙嗪（chlorpromazine）。凱特甚至說，實際上應用於精神分裂症治療的任何一個發現，原本**都不是**

為了治療精神分裂症而做的研究！我們一次次的看到，試圖控制科學的研究方向（亦即告訴科學家去解決一個實際問題的企圖）往往造成相反的結果：阻礙而不是促進了科學的進展。諷刺的是，積極想要科學只解決實際問題，而不要管「其他事」（基礎研究），往往是最不切實際及最短視的做法。

通往實際應用的道路，往往充滿艱辛與意外。德州西南大學醫學中心（University of Texas Southwestern Medical Center）有一組研究人員為了研究關節炎，找來一些患有關節炎的老鼠並進行染色體干預。出乎意料的，他們的老鼠竟然也患了腸炎（Fackelman, 1996），「我們沒想到會將老鼠弄成潰瘍性大腸炎，因而能為今後其他科學家研究人類疾病提供難得的有腸炎的動物樣本。」（Fackelman, 1996, p. 302）不管他們在治療關節炎上是否能有進展，但是他們看來卻為潰瘍性大腸炎的病因及治療做出巨大的貢獻。

簡言之，我們必須認識到，儘管有些研究是為了直接預測某一特殊情境而設計的，但是大多數的科學研究仍然意在驗證理論的基礎研究。要回答「這些研究結果怎麼應用到現實生活中？」這一問題，從事應用及基礎研究的工作者會有完全不同的答案。應用學者會說：「可以直接應用，只要實驗情境與將來應用的情境有相當程度的相似性就可以了。」因此，像樣本的隨機取樣性及實驗情境的代表性等問題，都會影響其結果的應用效果。然而，理論學者會回答，他們的發現不能直接應用於現實生活，他們的研究目的也不在於此，因此此時也不關心他的研究受試者與其他群體有多大的相似性，或實驗情境是否反映了真實生活中的情境等問題。但這是不是說這些研究結果就不能應

「但這不是真實生活啊！」

用於現實世界？當然不是。這些結果不能直接應用於現實世界，卻能直接應用於一個理論。而一個理論則會在將來與其他科學規律結合之後，用於解決某個現實問題。

這類透過理論而間接應用於現實生活的例子，在心理學的一些領域中比比皆是。例如，幾年前行動電話剛進入市場時，不少認知心理學家就意識到，如果人們開車時接聽電話很可能有交通上的隱憂，因此他們立刻預測使用行動電話將導致交通事故增多。他們推論的理由，倒不是因爲接聽電話時手會離開方向盤，而是駕駛的注意力將轉移到談話上。值得注意的是，心理學家提出這些擔憂遠遠早於真正用行動電話做的實驗研究（參見 Strayer & Johnston, 2001）。他們是用在心理學裡存在了幾十年、關於注意力有限能力的**理論**（如 Broadbent, 1958; Kahneman, 1973）而做出這樣的推測，開車時使用行動電話，正好可以用上這個經過無數次實驗驗證的理論來預測可能造成的危害。實際上，後來使用行動電話做的研究，也證實心理學家關於注意力理論的假設：使用行動電話的確可能引起交通事故（Redelmeier & Tibshirani, 1997, 2001）。

默克（1983）提到的一個例子，可以說明心理學中透過實驗驗證理論的理念以及間接應用性的本質。一九三〇年代，赫奇（Selig Hecht）在《普通實驗心理手册》（*Handbook of General Experimental Psychology*）中發表了對視覺敏感度（簡稱視敏度）的一系列研究（Murchison, 1934）。這些研究中談到適應黑暗現象。當我們走進黑暗的電影院時，會出現暫時的「失明」，但如果你在座位上待一會兒，會察覺到椅子及人，然後其他物體也都慢慢看得見了。再仔細留意，你會發現這一視敏度不斷升高的

現象會持續幾分鐘之久。

這種現象稱爲**暗適應**（dark adaptation），它分爲兩個階段：首先是剛剛進入一個暗黑房間時，視敏度迅速但小幅度的升高；在這之後，視敏度會大幅升高。赫奇將視敏度在這兩個時段不同的升高曲線，與視網膜上的兩種不同感受細胞聯結起來。其中視錐細胞集中分佈在中央凹（fovea，光線在視網膜匯聚的部位）的中心，它們對紅光格外敏感；視棒細胞位於中央凹的周邊，分佈不集中，對紅光也不那麼敏感。赫奇根據這些有關視網膜結構的知識提出一個理論：暗適應最初的階段（即視敏度迅速小幅度升高的階段）是視錐細胞的適應作用所引起，而第二階段（即視敏度的大幅升高）則取決於視棒細胞的適應作用。

默克（1983）提醒我們注意，赫奇的實驗是在完全非自然的情境中進行的。受試者（當然不是隨機取樣的）關在一間暗室裡，根據他們是否察覺到面前閃爍呈現的紅光點，回答「是的，我看到了」或是「沒有，我沒看到」。這種微弱的紅光點是我們在日常生活中根本不會理會的。然而，赫奇根本不關心他的研究結果是否可以推廣到在暗室中對一個紅光點回答「是」或「否」的老百姓，因此這種情境是否可以在現實生活中找到，對他來說無關緊要。他眞正感興趣的，是建立及驗證有關視覺系統結構的理論。暗適應這樣的視覺基本歷程的研究，正可以幫助他建立支持他的理論之證據。他並不關心他的實驗情境是否在現實中存在，而只關心這種情境能否有效地將他要研究的這一特殊的視覺歷程分離出來。

「但這不是真實生活啊！」

赫奇的研究發現之所以能取得普遍性，並不是因爲他的實驗情境性質是人工的或自然的，而是因爲他有能力建立一個有關視覺歷程的理論，這個理論可與許多視覺現象或運作相關聯，以致他可以利用它們驗證自己的理論。他的研究之所以能發現人類視覺系統中各個部分之間的功能關係，正是因爲實驗情境是受控制的並且是人造的。如果他的功能關係理論是對的，那麼將來的應用價值應該非常高，並且可用以解釋與發現與這一理論的實驗情境完全不同的情境中出現的行爲現象。

換句話說，赫奇的發現是透過它們對理論的驗證而間接地找到可以應用的地方。例如，透過赫奇提出的視覺系統理論，已經對治療夜盲症以及識別X光產生很大的幫助（Leibowitz, 1996; Mook, 1982）。更有意思的是，二次大戰期間，那些等待德軍夜間突襲的英國飛行員，都會戴上一副紅色的飛行眼鏡（因爲視棒細胞對紅光不敏感，可以停留在暗適應的狀態，參見 Mook, 1982）。從讓受試者判斷小紅點，到倫敦上空危險的空戰，這一鴻溝是透過理論跨越的，而不用將赫奇的實驗室重新設計成戰鬥機機艙的樣子。

心理學理論的應用

一旦了解大多數研究是爲了發展理論，而非對特殊情境進行預測，並且也知道大多數研究的發現是透過理論得到間接應用，而非直接應用於特殊情境，我們就可以接著問，心理學已經有多少透

過理論得到應用的例子。也就是說，我們想知道，有多少心理學理論曾通過普遍性的檢驗？

這方面，我們必須承認以往的記錄是極其糟糕的，但也必須清楚地意識到，這與心理學的多樣性息息相關。一些領域的研究確實在應用方面的進展甚微，然而，其他領域則已經取得令人矚目的成績，透過實驗已推導出許多具有解釋力及預測力的原理（參見 Byrnes, 2001; Gazzaniga, 2000; Kolb & Whishaw, 1990; Pinker, 1997; Scarborough & Sternberg, 1998; Sternberg, 1999; Wilson & Keil, 1999）。例如，即使在諮商心理、學校心理、臨床心理、心理治療等應用領域，都從理論導向的基礎研究中受益不少（Davidow & Levinson, 1993; Fagley, 1988; Garb & Schramke, 1996; Lilienfeld, Wood, & Garb, 2000; Mahrer, 2000; Siegert & Ward, 2002; Strupp, 1989）。

可以想像一下古典和操作條件反射的基本行爲原理，這些原理幾乎完全是在非常人工化的實驗室情境下，從鴿子、老鼠之類的非人類動物身上發現的，然而這些原理已經成功地應用於許多人類問題上，包括對兒童自閉症的治療，酗酒、肥胖的治療，精神病院病患的管理，恐懼症（phobia）的治療等。這還僅是一小部分而已。

這些原理之所以可以做這些應用，是因爲在實驗室的研究中，研究者可以精確地辨認外界刺激與行爲之間的關係，這是在自然情境下無法做到的，因爲自然情境裡許多行爲之間的關係可能會同時作用。至於用非人類的受試者，許多例子已經證明，從實驗動物行爲研究中得出的理論，是對人類行爲規律很好的初步仿眞，把人作爲研究對象時所發現的規律及定理，與用其他動物做實驗的結

「但這不是真實生活啊！」

果非常相似。其實，這一發現在如今應該不足爲怪的了，因爲人類疾病治療的每一步進展，現在幾乎都建築在動物研究蒐集的資料上（Kalat, 1995）。例如行爲治療、減緩壓力、心理治療、受傷或殘障人士的康復、老化對記憶的影響、對有神經肌肉障礙患者的輔助、藥物對胎兒發育的影響、藥物濫用、失憶、慢性疼痛的治療等，都是動物實驗推進的應用實例（Domjan & Purdy, 1995; Feeney, 1987; Kalat, 1995; Miller, 1985b）。最近，一個在狗身上進行的研究，對認識人類焦慮症背後的原理即將做出突破性的進展了（Groopman, 1999）。

實際上，「這不是眞實生活」的批評已經導致人們對動物實驗的刁難，而這種刁難往往是出於政治的原因。例如，那些爲重度污染企業效勞的政客，總是無視於在動物身上進行的有關污染是致癌的一個潛在因素的研究評估報告。他們宣稱動物研究不可信，不能推廣到人類身上，妄圖爲製造污染的企業開脫罪名。然而實際上，一九八八年路易斯安那州的一組科學家在對二十三種致癌物（包括苯、石棉等）的研究中發現，由動物實驗估計的死亡率，與人類流行病學研究取得的資料十分相近（Finkel, 1996）。

心理學家有關知覺歷程的研究，在理論上也取得了令人矚目的進展。他們從實驗中推導出的規律和理論，已經應用於解決各種不同的實際問題，例如雷達監視、街道照明和飛機駕駛座艙的設計（Nickerson, 1992; Swets, Dawes, & Monahan, 2000; Wickens, 1992）。同樣的，認知心理學家在研究人們處理記憶訊息的方式上，也累積了大量可重複的研究結果。雖然作爲心理學中最年輕的理論，認知心理

學的應用範圍還不能和其他老牌理論相提並論，但其可能性已經在工業、教育等情境中被檢驗。例如，老化對認知運作的影響已經累積不少知識（Rabbitt, 1993; Salthouse, 1996; Salthouse & Miles, 2002），而這些新知識，有可能直接幫助我們設計對認知喪失者進行補償訓練的方案（Dixon & Backman, 1995; Poon, Rubin, & Wilson, 1989）。

心理學家對判斷與決策的研究，已經應用於醫療、教育和經濟等領域的決策制定（Belsky & Gilovich, 1999; Davidow & Levinson, 1993; Dowie & Elstein, 1988; Fridson, 1993; Gilovich, Griffin, & Kahneman, 2002; Kahneman & Tversky, 2000; Swets, Dawes, & Monahan, 2000; Zweig, 2001）。一個令人興奮的新進展是，認知心理學在司法系統的作用逐漸受到重視，例如，資訊蒐集時記憶的問題、證詞的評估以及如何判決等方面，都爲檢驗認知心理學理論提供很好的機會（Bornstein & Rajki, 1994; Kassin, Tubb, Hosch, & Memon, 2001; Koehler, 1993; Kuhn, Weinstock, & Flaton, 1994; Loftus, 1993; Wrightsman, 1999）。另外，從一九八〇年代開始出現的許多閱讀教學的理論與實踐，也逐漸受到認知心理學的廣泛影響（Ehri, Nunes, Stahl, & Willows, 2001; Pressley, 2002; Rayner, Foorman, Perfetti, Pesetsky, & Seidenberg, 2001, 2002; Stanovich, 2000）。最後，在大眾關注的焦點問題上，心理學家也給了相當重要的科學依據，例如，如何對待兒童在法律程序中提供的證詞（Ceci & Bruck, 1993a, 1993b, 1995; Ceci & Hembrooke, 1998）以及受虐兒童「恢復」的記憶是否準確（Bremner, Shobe, & Kihlstrom, 2000; Clancy, Schacter, McNally, & Pitman, 2000; Loftus, 1994, 1997; Pezdek & Banks, 1996; Pezdek & Hodge, 1999; Spanos, 1996）。

「但這不是真實生活啊！」

正如前面提到的，認知心理學中的研究成果已經通過可重驗性這一基本檢驗。許多有關資訊加工過程的基本原理，已經得到全世界許多實驗的驗證和支持。但是很多人還是不了解，如果密西根大學的一位心理學家有某個重要的發現，那麼同類實驗很快就會在史丹佛、明尼蘇達、俄亥俄、劍橋、耶魯、多倫多等多所大學以及其他地方進行重驗。透過這種重驗，我們很快就能知道這一發現到底是密西根大學受試者的特殊性所致，還是研究的實驗情境所致。

當然，並非所有心理學研究結果都能成功的重驗（參見 Carroll & Nelson, 1993）。恰恰相反，重驗失敗不但有，而且從失敗中得到教訓和啓迪往往比成功的重驗多。然而在認知心理學研究中，因受試者的特殊性而造成的重驗失敗倒是很罕見。實際上，重驗失敗的原因，大多是實驗刺激與方法上的細微差異造成的。因此，失敗後，透過仔細檢查要在什麼實驗情境下才能促發某行爲現象，科學家可以從中更準確的了解這一現象，從而建立一個有關它如何產生的更精密理論。

或者有人會問，既然一些研究結果有重驗失敗的事例，還能應用嗎？又或者有人要問，當科學家之間還不能對理論的所有細節達成共識時，應用這種尙未定論的知識及理論是否太快了一點？這類的顧慮，在心理學知識的應用中經常可見。這是由於人們沒有意識到，其他科學其實也是在發現和理論還沒有完全確立之前就開始應用了。當然，我們在第二章就清楚地講過，任何一門科學都要經過不斷修正。如果我們在應用科學的研究成果之前一定要絕對地確定正確性，那麼什麼應用都不會發生。所以應用科學家會盡最大的努力應用最準確的現有知識，但是，同時也意識到該知識還是

有可能不正確。

◎「大二生」的問題

許多人很關注心理學研究成果的「代表性」問題。他們的質疑主要集中在研究中的受試者，而非實驗設計的細節。然而，在心理學的許多領域，例如基本感覺及認知歷程的研究中，存在的爭議多數是由實驗設計引起的。在基本的資訊加工、腦的功能結構以及視覺系統的特徵等方面，無論受試者是蒙大拿州人還是佛羅里達州人（甚或是阿根廷人），研究結果都相似，而且這些結果與受試者的父母是修補匠、裁縫還是教授都沒有任何關係。

所有科學都會假設一些因素對實驗結果不會造成什麼影響。生物學家一般不會擔心，培養細菌的器皿在厚度上的細微差異會對培養的細菌產生什麼明顯的影響。當然，這些差異有可能造成一定的影響，科學的假設不是絕對的，但是生物學家必須把精力放在他們認為更有可能是影響因素的變數上。這正是為什麼，赫奇假設暗適應與人們的宗教信仰無關，所以他不問受試者信基督教還是天主教。

在這裡，我們還面臨一個俗稱**大二生問題**，它是指參與心理學研究的受試者中，有極大比例是大學二年級生。由此問題引發了對心理學研究結果是否能推廣到其他人群的質疑。心理學家對此很關注，因為它在某些研究領域中的確是個嚴重的問題。然而，我們一定要只從某一個角度看這個問

「但這不是真實生活啊！」

題，不要泛化問題；也就是說，並不是所有研究都會因受試者是大二生而出問題。更何況，心理學家對這一批評有幾個合理的辯解。這裡提出三點：

1. 大二生的批評並不會令研究結果不正確，只是促使我們展開更多研究，以便對理論的可推廣性進行評估。由於我們先前蒐集了大二生的資料，即使從其他人群中獲得相反的資料，而必須對理論作出相應的調整，也只會使理論變得更精確，不會完全否定它（Gage,1996）。即使在比較極端的情況下，重驗結果完全失敗，充其量也只能說，建立在大二生資料上的理論不夠全面，不能說一定是錯的。
2. 在心理學的許多領域中，大二生並不構成問題。因爲在基礎理論（如視覺系統）的研究中，那些歷程是非常基本的，沒有人會認爲它們的基本組成會受到樣本的人口統計學特徵之影響。
3. 研究在不同的地區取得一致的結果，大則能增加理論的地域普遍性，小則可以增加在社會經濟、家庭、早期教育經驗等因素的普遍性。例如，與五十年前只有極小數精英才能讀大學的大二生相比，現在的大二生來自各式各樣的社會經濟背景，這與現在有許多的大學都已經平民化有關。

然而，在心理學某些領域中，不承認有大二生問題是不明智的。不過，心理學家正努力糾正此一問題。例如，基於研究的內容，發展心理學家必須非常關注這個問題。每年都會有數百名這方面

的研究者，將眾多用大學生做受試者得出的理論及發現，在其他不同年齡的受試者身上進行重驗。

從不同年齡群體得到的結果，與大學生的結果不會總是相同。對發展心理學家而言，如果老是得到完全相同的結果，就太沒意思了，因爲他們的用意是要將年齡因素加入心理學各個理論中，以顯示年齡的重要性。這一領域的研究也確保了心理學的宏大理論，不是只建立在從大學生那裡蒐集的、微薄的資料基礎上。

發展心理學家也進行一些跨文化研究，以驗證只根據北美兒童得到並建立的成長發展規律，是否具有全球普遍性。例如，史蒂文森等人（Stevenson et al., 1985）透過對中國、日本和美國孩子的大量認知任務實驗得出：「三種文化下的孩子在執行這些任務時，所反映出的認知能力結構是十分相似的。」（p. 727）其他跨種族、跨文化的比較也驗證了史蒂文森等人的發現（McBride-Chang & Kail, 2002; Rowe, Vazsonyi, & Flannery, 1994）。在其他領域，也有許多跨文化比較的研究顯示相似的趨勢（如 Day & Rounds, 1998; Krull et al., 1999; Rozin, Lowery, Imada, & Haidt, 1999）。當然，也有不少跨文化研究並不能成功地重驗用美國大學生所得到的研究結果（如 Menon, Morris, Chiu,& Hong, 1999; Nisbett, Peng, Choi, & Norenzayan, 2001; Peng & Nisbett,1999）。但當不一致出現時，這些研究結果仍然對某些理論及結果受文化環境限制這一特點提供重要的資訊（Chang, 1996; Choi, Nisbett, & Norenzayan, 1999; Yates, Lee, & Shinotsuka, 1996）。

教育心理學家也注意到大二生問題。例如，他們與發展心理學家及其他教育問題研究者合作，

「但這不是真實生活啊！」

共同制定一種測量基本認知能力的測驗，用於預測學生未來的學業成績，例如閱讀掌握率。結果有中等程度的準確度，而且這一準確度不受學校所在區域、社經水準以及種族等因素的影響。

對大二生及代表性問題的批評，多是針對社會心理學而來的，因爲這一領域經常使用大學生樣本，透過實驗室研究，以發展有關社交、群體行爲以及社會情境中的資訊加工過程等方面的理論（Kunda, 1999）。然而，即使是在這個領域，也有許多證據顯示，實驗室裡發展出來的關係及理論，還是能夠預測許多不同情境中不同類型的人之行爲。

例如，幾年前，威斯康辛大學（University of Wisconsin）的心理學家伯克維茲（Leonard Berkowitz）提出所謂的武器效應。它是指只要在個體的環境中出現武器，就會增加個體的侵犯行爲。這一發現最初是在實驗室中進行的，可說是不具代表性的典型實驗，因此受到猛烈的抨擊，認爲研究結果是在特別設計的情境中誘導出來的行爲。然而，無論研究者用的是什麼工具來測量侵犯行爲，無論是在歐洲還是在美國，無論受試者是兒童還是成人，無論是在實驗室之內還是在現實生活的情境——參與研究的受試者根本沒有意識到自己在做實驗，針對這個理論所做的研究都不能推翻它（Berkowitz & Donnerstein, 1982; Turner, Simons, Berkowitz, & Frodi, 1977）。有些研究者甚至將武器效應背後的認知機制分離出來，認爲是一個語義記憶被（武器的出現）自動啓動的過程（參見 Anderson, Benjamin, & Bartholow, 1998）。這個效應強烈到令安德森等人（1998）把研究報告的題目定爲：「是手槍自己扣動了扳機嗎？」

認知、社會以及臨床心理學家，都對人類決策行爲展開許多研究。最初，大多數的研究是在實驗室中進行的，並多用大學生作爲受試者，讓他們完成一些極其人爲化的任務。然而，許多從這些研究中得到的決策行爲原理，已經在大量的非實驗情境中成功地重驗，包括銀行家對股價的預測、紙牌賭博遊戲、精神病醫師對病患行爲的預測、經濟預測、軍事情報分析、美式足球聯盟（NFL）比賽的下注、技師對維修時間的估計、房地產經紀人對房價的估計、商務決策以及醫師的診斷等等（Bazerman, 1999, 2001; Braun & Yaniv, 1992; Davis & Holt, 1993; Goldstein & Hogarth, 1997; Hammond, 1996; Heath et al., 1994; Tassoni, 1996; Wagenaar, 1988）。這些原理，還在相當實用的個人理財諮詢領域中得到應用（Bazerman, 1999, 2001; Belsky & Gilovich, 1999; Kahneman & Tversky, 2000; Thaler, 1992; Zweig, 1998, 2001, 2002）。

博恩鮑姆（M. H. Birnbaum, 1999）預計，網際網路將爲心理學的大二生問題提供一個處理辦法。他曾在實驗室中做了有關決策問題的一系列實驗，並在網際網路上徵求一批志願者。結果發現在實驗室中用大二生的研究結果，與後來在網上徵集的受試者的結果相同，而後一批受試者的來源比前者廣泛得多——包括來自四十四個國家的一、二二四名志願參與者（同時參見 McGraw, Tew, & Williams, 2000）。

僅從這少數幾個例子，我們已經可以看到心理學研究結果的穩定性及普遍性經常被人低估（參見 Cheng, 2001; Gage, 1996; Lipsey & Wilson, 1993; Rosenthal, 1990）。安德森、林賽和布希曼（1999）曾報告

他們對實驗室研究與田野研究結果的比較研究。在對許多不同課題——如侵犯行爲、領導行爲和憂鬱症等的大量研究——進行考察之後，他們得出一個相當一致的結論：不同的研究情境並沒有帶來不同的結論，反而總是帶來相同的結論。

◎透視眞實生活與大二生問題

在本章我們提到幾個焦點問題，在這裡回顧一下，以便弄清楚哪些是我們說過的，哪些還沒說。我們談到，由於人們對科學研究的原則存在一些誤解，以致對心理學研究的人爲性經常提出批評。我們也談到，爲什麼人們會關注心理學家沒有在所有研究中使用隨機取樣的樣本，以及爲什麼這種擔憂沒有必要。最後，我們談到一個較爲合理的批評：大二生問題，但這個問題也常常過分誇大，尤其是那些對心理學研究的多樣性和廣泛性不熟悉的人（詳見第一章）。

儘管如此，心理學家確實應該謹慎處理他們的實驗結論，不要過分依賴一種方法，尤其是同一類受試者群體。在下一章，我們將對這一點進行討論。事實上，心理學中確實有一些領域深受大二生問題之苦（Wintre, North, & Sugar, 2001）；而作爲這個問題的解藥的跨文化心理學，則尚未發展成熟。然而，我們對這個問題可以緩解仍然抱有很大的希望，因爲在心理學研究者的行列中，自我批判是非常受到重視的（參見第十二章，Anderson & Gunderson, 1991; Dawes, 1994; Heinsman & Shadish, 1996; Kimble, 1999; Proctor & Capaldi, 2001; Redding, 1998; Wintre, North, & Sugar, 2001）。實際上，已經有許多著名

的心理學家以批判這個領域爲專業（Leary, 2001; Robinson, 2001）。每年各類科學雜誌上，也都會出現就心理學研究方法的漏洞或大二生問題提出警告；尤其是有關大二生問題，心理學界已經非常關注，想必沒有任何一位心理學家未知覺到這一問題的存在。所以，我們固然不應忽視，但也不要將之過分泛化。

摘要

有些心理學研究是應用性的，目的是將研究結果直接應用於特定的情境中。在這類應用性研究中，結果是直接推廣到現實中的自然情境裡，因此樣本的隨機取樣性與情境的代表性相當重要。然而，心理學的大多數研究卻不屬於這一類。它們大多是基礎研究，用來驗證人類行爲背後機制的理論。在這類基礎研究中，受試者是不是隨機取樣及研究情境是不是有代表性，並不是特別重要，因爲這些研究關注的重點，是驗證一個理論之預測的普遍性。事實上，在驗證理論的基礎研究中，正如上一章說明的，人爲的情境是故意設置的，爲的是將研究中的關鍵變數分離出來，並且用之控制那些無關緊要的變數。從這個角度視之，心理學實驗的「不是眞實生活」，恰好是它的長處，而非短處。

「但這不是真實生活啊！」

第八章　避開愛因斯坦併發症

——聚集證據的重要性

「生物學實驗披露生命的奧祕」、「心智控制獲重大突破」、「加州科學家發現延緩死亡之祕訣」——這些形形色色、充斥於小報頭條的所謂「突破性」新聞，可謂屢見不鮮。正因爲這些新聞通常源自那些最不負責的媒體，難怪大多數科學家都建議大衆，要用懷疑的態度看待這些新聞。不過，本章的目的，不僅僅是告誡大衆有關這些經由誇大而散播的錯誤資訊，或是警告讀者在評價有關科學進展的報告時，必須審愼地查核消息來源。這一章中，我們還想引入一個比前幾章講述的更複雜的科學進展觀。爲此，我們將詳細闡述第一章介紹過的系統實徵主義和公開的知識這兩個概念。

媒體常用的突破性大標題，其實是誤導了大衆對心理學和其他科學的認識。一個特別典型的曲解就是，它們誘使大衆認爲透過一個關鍵性的實驗，就能解決某一科學研究領域的所有問題；或者說，科學的進展可以由具有批判性的某個單一新見解一蹴而成，而這一新見解能夠完全推翻先前衆

多研究者累積的所有知識。這一廣為流行的科學進展觀，倒是與新聞媒體的操作方式十分吻合。就媒體而言，歷史往往是用一些支離破碎、互不關聯的事件連串起來的。此外，此一科學觀也與好萊塢娛樂業的作業方程式相近，好萊塢的電影劇本必須有一個開始及一個令人滿意的結尾，令人滿意是指不能留有任何的懸疑及不確定因素。然而，這一觀點實際上只是對科學進展相當粗糙的描繪。說得更嚴重些，它還導致人們對科學進程的曲解，並減弱他們在某個問題上評估科學知識的能力。在本章中，我們將討論科學研究的兩個原則——關聯原則（connectivity principle）和證據聚集（converging evidence）原則。與前面所述突破式模型的進展相比，這兩個原則能夠更準確的描繪科學進展。

關聯原則

儘管我們不認同科學進展取決於「大躍進式」或關鍵性實驗的觀點，但並不是說過去從來沒有過關鍵性的實驗和理論上的大躍進。相反地，在科學史上一些最著名的實例，都證明這一類的實驗及躍進的確存在。其中最有名的，可能要數愛因斯坦提出相對論的發展過程。這個例子中，一系列讓人驚歎的理論洞察力，導致人們最終重新構想諸如時間、空間和物質等基本概念。

然而，正是愛因斯坦取得的這一輝煌成就，成為大衆心目中關於科學進展的主導模型。這一模型之所以一直延續下來，主要是因為它與媒體慣常用來報導新聞的隱含「腳本」高度吻合。人類歷

史上還沒有哪個觀念遭遇像相對論那樣的麻煩：那麼多的胡言亂語，那麼多的不實推論都緊隨其後（沒有，愛因斯坦並沒有證明「一切東西都是相對的」——參見 Holton, 1996）。當然，這裡我們的目的並非糾正這些謬誤，但是其中之一倒可以幫助我們在稍後討論如何評價心理學理論時，帶來一些啓迪。

愛因斯坦理論引發人們對物理世界最基本觀念的重新認識，正是這樣的基礎及重要性，許多通俗讀物中常常視之爲類似藝術中的觀念轉變（一位二流詩人被重新評價成一個詩壇天才、一種藝術流派被宣告死亡等等）。然而，此一比較恰恰忽略了藝術與科學在觀念轉變中存在的基本差異。科學研究中的觀念轉換遵從關聯原則，這一原則在藝術中並不存在，或至少是極爲少見的（參見 Bronowski, 1956, 1977; Popper, 1972）。

一個新的科學理論，必須與先前已建立的實徵資料連接，只有在解釋新現象的同時又能與原有的事實證據相容，這一理論才會眞的被認爲是一個進展。新的理論可以用與先前理論截然不同的觀點解釋舊證據，但必須解釋得通。這一要求確保了科學能夠不斷地累積、向前發展。除非一個理論的解釋能力所適用的領域能夠大大拓寬，否則不會有眞正的進步。如果一個新理論能夠解釋新觀察到的現象，卻無法與大部分的舊事實相容，那麼人們不會承認它是對舊理論的全面更新，也因此不會立刻用它替代舊理論。相反地，在這種情況下，舊理論和新理論會在觀念的角力場上繼續搏殺，直到有一個可以整合兩者的新理論出現，屆時這兩個角力者都被淘汰出局。

儘管愛因斯坦的理論帶來令人歎爲觀止的全新觀念（運動中的時鐘走得較慢，數量隨速度增加），仍

然遵從關聯原則。它雖然捨棄了牛頓式機械力學，但沒有否認或捨棄牛頓理論的基礎，即有關物質運動的實徵證據。恰恰相反，低速運動時，兩個理論做出基本相同的預測。愛因斯坦體系的高明之處在於：它能解釋與牛頓力學不相容的衆多新發現的、有時甚至令人意想不到的現象。我們因此看到，即使是愛因斯坦的理論，這個科學史上最驚人、最基本的概念重構，實際上都遵從關聯原則。

◎消費者的準則：留心是否違背了關聯原則

科學進展的大躍進式模型——我們姑且稱之爲愛因斯坦併發症——經常將我們引入歧途，因爲它讓我們覺得新發現必定是要違背關聯原則。這個觀念非常危險，因爲一旦放棄了關聯原則，主要的受益者將是那些僞科學和虛假理論的傳播者。這些理論的動人之處和大衆魅力部分是來自：它們常常披著令人驚歎的新外衣。「你看，相對論不是也在當時被認爲是全新的嗎？」這一辯詞已經變成一種策略，試圖將新奇轉化爲一種優勢。當然，一旦僞科學家試圖打入某些科學領域，這些領域先前所累積的大量資料就成了主要障礙。實際上，這一障礙也是微不足道的，因爲僞科學家還有兩個強大的法寶。我們在第二章曾經討論過其中的一個策略，即透過使自己的理論在解釋先前資料時，永遠無法證僞。當然，也令其變得毫無用處了。

否認先前實徵證據的第二個法寶是宣稱這些證據不相關。這種否定通常會著重強調其新理論的突破性。因此，常常看到類似「對現實的新觀點」和「全新的突破」之類的用語。但實際上，花招

還在後面。因爲接著，這個理論會被認爲是如此嶄新，以致使依據其他相關理論所獲得的實驗證據都被認爲是無關的，而只考慮根據這一新理論的概念框架所蒐集的支持性資料。這就徹底打破了關聯原則。很明顯的，由於理論是全新的，以至於他們可以理直氣壯地說：與之相關聯的實徵證據尚不存在。這樣一切都沒有問題了：僞科學擁有一片能夠蓬勃成長的沃土。那些舊的、「不相關的」資料都沒了，而新的、相關的資料尚未出現。正是由於愛因斯坦併發症讓人忘記了關聯原則，才使得這種詭計得逞。然而頗具諷刺意味的，恰恰是愛因斯坦的理論本身，論證了關聯原則的重要性。

加州大學（University of California）的古生物學家帕典（Kevin Padian）舉出另一個例子，說明當關聯原則的重要性被忽視時，大衆將如何曲解科學的本質。堪薩斯市教育委員會曾決定把演化論排除在學生的必修課程之外，帕典針對此事指出：「可見對科學如何進行整合人們完全缺乏理解……那種認爲可以把一部分科學——特別是像演化論這樣一個可以把整個生物學組織整合起來的主要理論——捆起來丟掉，而不會產生任何壞影響的想法，實在是太荒唐了。」（Carpenter, 1999, p. 117）。生物哲學家魯斯（Michael Ruse, 1999）注意到演化論與許多獨立、相關不大的科學領域——古生物學、胚胎學、生物形態學、生物地理學、神經科學等——都具有關聯性。與此類似，夏默（M. Shermer, 1997）指出「假設宇宙和地球只有大約一萬年老，那麼當代科學的知識，例如宇宙論、天文學、物理學、化學、地理學、古生物學、古人類學及早期人類歷史都是殘缺的。」（p. 143）知名科普作家和古生物學家古爾德表示：「失去演化論的生物學教學，就如同教授英語但不教文法。」（Wright,

1999, p. 56）

魯斯（1999）以一個例子說明達爾文如何善用關聯原則。當他的新理論與其他學科之間沒有必須有的相容性時，達爾文就斷然放棄。當時，達爾文想尋找一種遺傳機制與他的自然淘汰理論匹配，他試圖建立一個稱爲泛生論（pangenesis）的理論，這個理論認爲「身體的各部位都會給出一些很小的芽球，這些芽球在體內循環，並在性器官聚集，從而將之傳到下一代。」（p. 64）該理論的問題之一是它與細胞理論不符。另外，達爾文也無法解釋這些芽球是如何傳送的，因爲血液輸送實驗已經證明它們無法透過血液傳送。由於這個以及其他種種原因，泛生論最終「因爲與生物學裡的其他領域不相容」（p. 64），而在科學陣營中雲消煙散。

心理學也一樣。哲學家邦吉（Mario Bunge, 1983）曾提出，如果認知心理學一開始就否認經典和操作性條件反射之存在，它絕不可能在心理學領域中立足，因爲它就與行爲科學中的其他知識無法建立相關了。回想我們在第六章對溝通輔助的討論，我們之所以懷疑它治療語言障礙的功能，是因爲它違背了關聯原則。如果它是正確的，即意味著我們要推翻所有神經、遺傳及認知心理學等不同性質的科學領域中累積的基本原理。這一治療的理論假設與科學中的其他知識沒有任何關聯。反對演化論的神創論也是如此，它的問題也在於它與科學中的其他知識毫無相關，無論是生物學、地質學、生態學、化學或遺傳學。演化論則不同，正如生物學家多布冉斯基（1973）在一篇著名的文章中說的：沒有演化論，生物學將沒有任何價值。演化論與其他科學有著密不可分的關聯。

◎大躍進模式 VS.逐漸整合式模式

將愛因斯坦式的革命當作科學研究的典型，會誘導我們認爲科學進步依賴於許多的大躍進。這方面的愛因斯坦併發症並非完全憑空捏造的。在許多科學進展中，的確存在著在概念理解上的大躍進。問題在於，人們總是傾向於把這些例子泛化成科學進步**理所應當**的模式。實際上，許多科學領域所取得的進展，並非依靠單一的突破，而是靠一系列難以名狀的、停頓及開始的交互更迭。

一般人通常察覺不到在科學工作中涉及的模糊性，科學實驗很少完全解決某個研究問題，亦即在支持一個理論的同時，可以排除所有其他可能解釋同一現象的理論。新理論極少會很明顯地優於所有先前存在並與之競爭的概念體系，某一特定問題的解決通常不像科學電影射影的那樣，是靠某個關鍵實驗一蹴而成，而是要等到科學界逐漸開始產生共識，認爲支持一種理論的證據比支持其他任何理論的強很多時。科學家用來評估的證據並非源於單一、設計完美的某個實驗；相反地，科學家常常必須對幾十篇實驗論文的結果進行總體評估，這些實驗每一個都有缺點，又都提供了解決問題的部分答案。這一科學進程的更迭模式不常爲人知，是因爲愛因斯坦併發症在人們的心目中造成一個印象：所有科學都等同於物理學，而物理學是出現大躍進式科學進展模式最多的一門學科。

試想一下，近幾百年來，遺傳及分子生物學的突飛猛進，並不是靠一個愛因斯坦式的天才，在一個關鍵時刻把事情搞清楚的。生物學的現代整合，是建立在無數次各有缺陷的實驗研究所帶來的

洞見之上；這些進展並不是靠對一個重大概念上的創新的即時認可，而是在幾種得到支持的可能解釋之間長期、反覆的拉鋸中逐漸形成的。經過十年沒有結論的實驗、無數的理論推論、辯論與批評，科學家才改變了他們對基因到底是由蛋白質還是由核酸組成的觀點。統一的觀點雖然改變了，但絕對不是在一次大躍進中完成的。原子核的發現者盧瑟福（Ernest Rutherford）在強調關聯原則的重要性時，也指出「科學家依賴的，不是某一個人的思想，而是成千上個人的智慧結晶。」（Holton & Roller, 1958, p. 166）

盧瑟福的觀點點出另一個區分科學和僞科學知識的方法。科學是一種累積式的運作，它遵從的是關聯原則，特性之一就是有許多人參與，而我們是以這些人能幫助我們理解自然界之程度決定他們的貢獻。沒有任何單一的個人，可以僅憑其特殊的地位而控制科學的進程。科學不允許有只有少數人才能享用之特殊知識的存在。在第一章，我們實際上已經討論了科學的這一公開性。相反地，僞科學則常宣稱某些權威和研究者有一種「特殊」的管道，使他們比別人更能找到眞理。

我們在這裡講述了兩個觀點，它們提供了幫助我們理解心理學這門學科所需的背景。首先，科學中不存在任何設計得十全十美的實驗，任何一個實驗對它資料的解釋都存在一定的模糊性。科學家通常不會等到完美或關鍵的實驗出現後才評估一個理論，而是將評估建立在大量有些缺陷的實驗所呈現的整體趨勢上。其次，許多學科儘管沒有像愛因斯坦那樣的人物出現，仍然向前發展。它們的進展是以動靜更迭的方式，而不是通過愛因斯坦式的偉大整合、飛躍的方式進行的。這是因爲心

理學和許多其他科學一樣，是由眾多不斷增長的知識碎片聚集而成的，缺乏一個整合的主題。

證據聚集：在缺陷中進步

先前的討論引出在心理學中做證據評價時至關重要的另一個原則。這個原則常稱爲**證據聚集原則**（**或操作聚集原則**），科學家和應用科學知識者經常需要判斷，大量的研究證據究竟在說什麼。在這種情況下，證據聚集原則成爲一個重要的工具。對使用科學資訊的外行人來說，證據聚集原則也很有價值，而且在評估心理學的療效特別有用。儘管對證據聚集原則展開全面技術性討論會偏離本書的主題，有關這個概念的實踐應用卻是很容易理解的。下面，讓我們探討一下這個原則的兩種表達途徑，一種稱爲缺陷實驗的邏輯，另一種稱爲理論檢驗。

從極端觀點看，一個實驗存在無數個出錯的可能性（或用專業術語來說，就是變得**混淆不清**）。然而在大多數情況下，不會存在數量很多的關鍵性混淆因素。對一個特定問題，較有經驗的科學家通常很清楚大多數的關鍵因素是什麼。所以當考察實驗證據時，科學家通常知道每個具體實驗的關鍵性缺陷是什麼。於是，證據聚集原則提示我們去審察相關的研究文獻中呈現的缺陷類型，因爲這一類型的性質要麼支持、要麼削弱我們想得到的結論。

例如，假設從一系列不同實驗得到的結果一致支持某個特定結論。既然這些實驗都有缺陷，我

們可以對這些研究存在缺陷的程度和特性進行評估。如果這些實驗都擁有相同的缺陷，將削弱我們對研究結論的信心。因爲人們會假設：研究結果的一致性，是由所有實驗共有的某個缺陷造成的。另一方面，如果所有實驗的缺陷各自不同，我們對實驗結果就比較有信心，因爲看上去，研究結果的一致性不至於是由某一污染因素而令所有實驗結果變得混淆不清。正如安德森與安德森（1996）指出的：「不同的方法似乎孕育出不同的假設。當建立在不同假設基礎上的某一概念假設能夠通過衆多基於不同假設的證僞檢驗時，我們可以說得到了一個有說服力的結論。」（p. 742）

每個實驗都能幫助我們校正其他實驗設計中存在的錯誤，同時它自身又在接受其他研究對其缺陷的檢驗中得以鞏固。大量的實驗，儘管各自多少都存在不同的缺點，實驗技術也各有優劣，但只要能得到近似的結果，我們就可以說實驗證據聚集在一起了。儘管沒有一個實驗設計十全十美，但我們還是得到一個強有力的研究結論。於是，證據聚集原則要求我們，必須將研究結論建立在一系列彼此略微不同的實驗所得的資料基礎上。這個原則使得我們可以做出強有力的結論，因爲這樣所得的結果一致性，比較不可能是由某個特定的實驗程序造成的。

證據聚集原則也可以用理論檢驗的方式加以論述。只有當一系列實驗一致性地支持某個理論，而又集體地排斥其他重要的競爭理論時，我們才說這一研究領域是高度聚集的。儘管沒有單個實驗可以排除所有可替代的其他解釋，然而如果把它們放在一起看，只想排除一兩個其他解釋的一系列實驗資料，呈現出某一個特別的形式，也可以讓我們下一個比較強有力的結論。

最後，引入證據聚集原則有助於我們驅除另一種錯誤觀念，這種觀念是由於我們在第二章討論可證僞性的時候，過分簡化的討論造成的。在那裡的討論似乎讓人感覺，當第一個與自己的理論不相符的證據出現時，這個理論就算是證僞了；然而，事實並非如此（Pigliucci, 2002）。正如理論是被聚集的證據所支持，它也是被聚集的研究結果所否定。

◎心理學中的證據聚集

強調聚集的重要性，原因在於心理學的結論通常是靠聚集的證據確定的。這一事實當然並非獨特或與衆不同（許多其他學科的結論，也都不是建立在具有決定性的單一實驗基礎上，而是靠對幾十個模糊實驗的整合）。但是我們有理由認爲，這種情況在心理學中尤爲突出，因爲心理學實驗通常鑒別原因的能力都比較低。也就是說，支持某一理論的資料，往往只排除一小撮可替代的可能解釋，同時保留許多其他可能的理論作爲候選。結果，只有蒐集大量研究資料後，才能得到較爲集中、因而較爲有力結論。

心理學實驗具有高度不確定因素，這點毫不令人驚訝。因爲它研究的問題涉及人類行爲，具有巨大的複雜性，如果心理學家能公開承認這點，並且致力於向大衆解釋這個事實所帶來的後果，人們可能會對心理學有更多的理解。心理學家應該承認，心理學是一門科學，並且正在發展之中，然而它的發展很緩慢，而且我們的結論往往來自相當痛苦的、日以繼夜的整合與爭論。對媒體經常宣

稱的所謂突破性進展，我們要永遠帶著懷疑的態度，這點在心理學領域尤爲重要。

在心理學這一領域，我們必須格外小心。例如，我們必須抗拒將一個周邊的實驗證據尚很模糊的心理學假設，視爲已經「證實」的理論。本書中好幾章都一再強調大家必須抱持這種懷疑態度，不要從相關中推論因果，以及拒絕接受見證敘述式的證據等，都是例子。與此同時，我們也不要因爲知識不完整或結論還有待進一步探索就妄自菲薄，開始懷疑心理學究竟能否產生強有力的結論。我們更不可因此就做出不理智的宣稱，認爲心理學永遠也無法成爲一門科學。從這個立場出發，證據聚集原則可以看作是前面提到的，不要過度解釋尚不成熟的研究結果這一警告的一個平衡鉈。儘管心理學研究有許多缺陷，但是不要氣餒，各式各樣研究的逐漸集中到一個方向，還是能夠讓我們得到合理、有力的結論。

觀察證據聚集原則最好的方法，就是檢驗存在問題的心理學領域，這些領域往往透過應用聚集原則而得到結論。讓我們再看一個例子，關於電視暴力節目是否增加兒童之攻擊行爲這一課題，能充分說明證據聚集原則的重要性。對這個問題，科學界的認識目前較爲一致：兒童觀看電視暴力節目，確實增加了攻擊行爲發生的機率；效應雖不是非常大，但確實存在。科學家對這結論的信心，是透過分析上百個不同的研究結果，而不是一個具有決定性的研究得到的（如 Anderson & Dill, 2000; Anderson, Huston, Schmitt, Linebarger, Linebarger, & Wright, 2001; Bushman & Anderson, 2002; Friedrich-Cofer & Huston, 1986; Huesmann & Eron, 1986; Paik & Comstock, 1994; Pearl, Bouthilet, & Lazar, 1982; Wood, Wong, &

Chachere, 1991）。這些研究的實驗設計、受試者群體及具體的技術都非常不同。我想現在大家一定清楚了，這些差異恰恰是這個領域的力量，而不是缺點之所在。

電視公司的老闆自然不願意看到自己充滿暴力的節目對兒童存在負面影響的證據。於是，他們就發起一場打擊這些研究在「誤導大眾」的宣傳攻勢，他們利用的正是大眾沒有認識到：研究結論是建立在許多研究的聚集證據上，而不是哪一個關鍵性、決定性的研究（因爲大衆正是愛因斯坦併發症的受害者）。電視公司不斷挑出**某個**研究加以批判，並且暗示透過論證每個研究都存在缺陷，就可以否定相關的總體結論。儘管社會科學研究者可以一一回應對這些研究的批判，但是大多數時候，研究者很樂意承認這些研究是有缺陷的。這一點是大衆沒有認識到的。這是因爲，本著證據聚集原則，研究者並不認爲承認某個特定研究的缺陷，就否定了關於電視暴力影響攻擊行爲所得到的一致性結論。這表示，在支持一個結論的衆多研究中，也許某些帶有這樣那樣的缺陷，但是另一些沒有這些缺陷的研究也得到同樣的結果。這些研究自身可能有問題，但另一些等於糾正了這些缺陷的研究，也得到類似的結果。

例如，有關電視節目與暴力行爲的關聯問題的早期研究揭露，兒童觀看暴力節目的數量與暴力行爲之間存在相關。當時就有人正確地指出，這種相關證據不能視爲因果結論，有可能是第三變數造成這一關聯，或攻擊行爲高的兒童傾向於看比較多的暴力節目（方向性問題）。

然而，科學界並非僅依賴上述相關證據就做出任何結論。我們也可以用許多比簡單測量兩變數

之間關係更複雜的相關分析技術，這些技術還允許我們做出一些探索性的因果推論（第五章中提到的偏相關就是其中之一）。這些技術之一，就是採用縱向的研究設計，在不同時間點測量相同的兩變數——此處是電視暴力和攻擊性。由這一設計所得的相關樣式，可以告訴我們兩者是否有因果關聯。這一類型研究已經有人做過，得到的結果表明：觀看暴力電視節目，的確增加人們在日後產生攻擊行為的可能性。

也許，這時有人又會質疑縱向相關技術具有爭議性。確實如此，但最重要的一點是，電視暴力和攻擊行為之間有因果關係的結論，並不完全依靠相關證據，不管是簡單相關或是像縱向相關技術這樣複雜的相關分析。實際上，已經有許多研究是用實驗法進行驗證工作的。在這些實驗中，大量電視暴力是被研究者操弄的，而不是僅僅對已有的節目（現實中自然的情境）進行暴力評估。在第六章我們討論過，在與其他變數受到實驗控制的同時只對一個變數加以操弄，就可以避免相關研究中對因果及第三變數問題的困擾。如果兩組其他變數完全相同的兒童表現出不同攻擊行為水準，而且兩組僅有的差異是一組看了更多暴力節目另一組沒有，我們就可以比較放心地推論：被操弄的變數（電視暴力）引致結果變數（攻擊行為）的改變。在大多數相關的實驗研究中都得到這一結論。

這些研究當然又引來另外一些批評，提出「這不是眞實生活」的抗議。前一章我們已經討論過這一抗議的論點，以及人們利用這些論點進行惡意的攻擊。無論如何，有關電視暴力的影響之研究結果，並不是只在一群古怪的孩子身上發現的；事實上，這一結果在美國不同地方及世界上不同國

家中都成功地得到重驗。在使用不同的實驗情境、不同的電視節目作爲研究對象時的各個實驗中，也都得到高度一致的結果。

重要的是，不僅在實驗室中，就連田野研究也大都得到相同的結論，一種稱爲**田野實驗**（field experiment）的研究設計也用來探討電視暴力—攻擊行爲這一課題。這類研究設計的發明提醒我們，避免慣性地認爲實驗設計和實驗情境之間存有某種必然的關聯。人們有時候認爲只有在實驗室裡才能操弄變數，只有在田野研究中才能探索相關。這個想法是不正確的。實驗室裡常常可以進行相關研究，而非實驗室情境下也常常可以操弄變數。在非實驗室情境下操弄變數進行田野實驗研究，儘管有時需要相當大的創意，卻在心理學中變得愈來愈普遍了。

當然，田野實驗研究自身也存在缺點，但是這些缺點往往正是其他研究設計的強項，所以可以互補。總而言之，電視暴力導致兒童攻擊行爲的可能性增加，支持這個觀點的證據，不是僅依賴於某一個或某一類特定的研究形式。

這個情況類似吸煙和肺癌的關係研究。煙草公司老闆常常試圖以這種方法迷惑大眾：香煙引起肺癌的結論，只存在於某些特定的研究中；然後他們再攻擊這些研究站不住腳。然而，大量證據都集中指向一個強有力的結論。來自若干不同類型的研究，資料都趨向聚集於一個結論，而且不會因爲單一實驗受批評而發生重大改變。

這裡很適合討論類似肺癌病因之類的**醫學**問題。在**醫學**診斷和治療中，許多決策都建立在不同

類型研究結果是否聚集得出一個結論的基礎上。例如，當由流行病學（可以說是一種涉及人類的田野研究，目的在尋求某一疾病與田野中的其他環境及地理因素的關聯）高度控制的動物實驗和人類受試者的臨床試驗等不同類型的研究之結果，都趨向聚集於一個結論時，醫學界才會對這一結論具有比較大的信心，認定這一結論是可靠的，而醫師才願意在這些證據的基礎上展開治療方案。

然而，上述這三類研究各有缺陷。流行病學研究通常是相關研究，變數之間存有虛假關聯的可能性很高；實驗研究雖然具有高度控制，但受試者往往是動物而非人類；臨床試驗雖然是在醫院環境下，將人類受試者置於一個眞實的治療環境中，又因此而有許多因素沒有受到控制，例如安慰劑效應以及治療小組對病患的治癒期望等。雖然各有問題，但是當不同方法所得的資料都緊密地聚集在一起時，醫學研究者就有理由肯定這個高度聚集的結論。就像心理學家可以用這一證據聚集原則幫助他們對諸如電視暴力對攻擊性行爲之影響的研究，做出強有力的結論一樣。

科學研究的共識

對電視暴力影響的評估是個典型的例子，告訴我們如何透過資料累積，最終回答心理學所提出的問題。特別是當涉及一些迫切需要解決的社會問題時，更需要牢記這些問題的答案總是不能信手拈來的，總要累積許多不同實驗的結果，絕不可能因爲出現一個突破性研究就一下子把問題完全解

決。用一個簡單的道理總結：在評估心理學的實徵證據時，心中要想的是**科學共識**，而不是**重大突破**；是**逐漸聚集**，而不是**飛躍進步**。

遵守「科學共識而非重大突破」原則，在圍繞著「幼童補償性教育計畫」的種種爭論中，起了很重要的作用。在一九六〇年代後期及七〇年代初期，就詹森（Lyndon B. Johnson）總統爲建立一個「優質社會」提出的一系列社會政策專案是否眞的有效，人們曾展開許多爭論。在爭論進行得如火如荼之際，大衆經常在報紙上看到這樣的頭條：「早期介入可將智商提高三十分」和「心智開啓失敗」等等。一個外行人如何面對如此矛盾的資訊呢？在這個例子中，「科學共識而非重大突破」原則，無疑可以提供一些幫助，因爲這一原則提醒我們，兩個報紙頭條所指的實徵結論可能都言之過早。實際上，也確實又花了研究者十年的時間，才對這一重要的社會問題找到科學研究的共識。

這一共識不是建立在單一的、關鍵性的研究基礎之上，而是建立在康乃爾大學一組專家的精心分析上（Lazar, Darlington, Murray, Royce, & Sniper, 1982）。進行分析時，這組研究者合併了一九六〇到七〇年代早期蒐集的，由十一個幼稚教育專案提供的成百上千個受試者資料。儘管單一的研究專案的結果有時難以解釋，一旦匯聚起來，整體結果卻還算清晰。早期介入的短期教育專案，並不一定會將智商提高三十分；同樣的，心智開啓和諸如此類的專案，也並非絕對失敗。早期教育介入專案確能對參與此專案的兒童之後期教育產生具體的、可靠的影響。這些兒童留級的比較少，送入特殊教育班較少，對學校和學業成績有比較正面的態度，並表現出持續的學業進步（參見 Lee, Brooks-Gunn,

Schnur, & Liaw, 1990; Ramey, 1999）。

加拿大心理學家摩爾（Timothy Moore, 1996）曾討論如何透過增加人們對證據聚集原則的敏感度，促使法庭審判可恰當的應用專家證詞。他特別討論到，當某個專家的證詞只是個人意見表達、並不能代表該領域專家的共識時所出現的問題。摩爾以一個叫猶大教士案例（Judas Priest trial）之專家證詞爲例。這個案例是關於兩個少年死於自殺的事故，他們的父母將一個名叫猶大教士的樂隊告上法庭，因爲他們認爲該樂隊的音樂隱含某些潛意識的訊息，教唆他們的兒子自殺。儘管專家證詞指出，當時科學在這一領域的共識是，沒有任何證據顯示潛意識訊息會產生導致人走向自殺的效果**（即使現在，這一共識仍然成立）**，然而該案的法官顯然受到某個堅信用心理動力學來解釋這一事故的學者之影響，沒有聽從科學界大多數人的共識。摩爾結論道：

> 但我們最終的分析結果是，誤導法庭的不是那些一眼就可望穿的偽科學，而是這位專家……他的觀點雖然極富想像力和邏輯性，卻與先前科學界對這個現象的理解大相徑庭。出色的履歷和出眾的職位，並不足以保證一個專家的觀點具有科學依據。任何一個專家，如果他的證詞獨樹一幟，風格奇特，且不為科學界較多數的同行接受，他就是沒有在法庭上真正發揮自己應發揮的作用。專家證詞在法庭的主要作用，是提供對這一領域不熟悉的人一些專業知識。（p. 38）

◎方法和聚集原則

聚集原則也告訴我們，心理學所有的研究領域都應使用多種方法研究同一個問題。由於不同類型的研究技術各有千秋，從各種方法得到的結果中求取平衡以對一個問題下結論，是比較妥當的。長久以來，心理學一直被批判爲過分依賴實驗室的實驗技術。這種批評也是心理學家相互爭論的話題，其有效性要看我們是在討論哪個特定的研究領域。但不可否認，心理學的各個領域近年來已經開始使用不同的研究方法了。例如社會心理學家，他們曾因爲過於依賴實驗方法而受到最多批判，如今都開始大量轉向富有想像力的田野設計，從而可以藉證據聚集原則支持其理論（Kunda, 1999）。

心理學家拉坦（Bibb Latane）和達利（John Darley）的工作提供一個很好例子。他們因研究所謂**袖手旁觀現象**而廣爲人知。這個現象是指，在緊急事件發生時，有些人不向受害者施助的現象。拉坦和達利（1970）發現，在許多緊急事件發生時，一個旁觀者會向受害者施助的可能性，會因爲有其他旁觀者在場而降低。

然而，兩位研究者清楚地知道，僅依靠觀察實驗室中的受試者遇到緊急事件時的反應下結論，就太令人置疑了。因爲在實驗室中，目睹緊急事件的受試者都是志願到實驗室參與實驗的，這些人可能比一般人更願意幫助別人。爲此，拉坦和達利設計了一個有趣的實驗，希望用另一個情境觀測這個現象。他們找到一家願意合作的賣酒商店，該商店同意假裝店裡發生盜竊事件。當收銀員跑到

店鋪後面爲某個「顧客」拿啤酒時，該「顧客」（實際上是研究者的共謀）拿了一箱啤酒走出商店大門。這一假盜竊案的上演，總是安排在收銀台前正好有一或兩位眞正顧客目睹時。收銀員回來後問這一或兩個顧客：「嗨，剛才在這裡的那個人哪裡去了?你看見他離開了嗎？」這樣，就給顧客一個機會告訴收銀員剛才發生的盜竊事件。與實驗結果一樣，當有另一個旁觀者在場的時候，亦即有兩個眞正顧客在場時，告訴收銀員盜竊案的行爲變少了。

並非只有社會心理學家企圖在不同的情境中重驗研究結果，認知心理學家也開始探索，如何將許多實驗結果向外推廣。例如，吉格任瑟（G. Gigerenzer, 1984）研究了頻率—效度效應的可推廣性。這個效應是指，一個不熟悉但看似有理的論斷，不管是眞是僞，僅僅經過不斷地重複，也會增加人們的相信程度。這個效應被多次成功地重驗，但僅限於實驗情境中，用大學生爲受試者（並且絕大多數在美國）。於是吉格任瑟做了一個在非實驗室情境下、用非大學生作受試者的研究。他在德國慕尼黑的許多成年非大學生的家中進行驗證，也發現有頻率—效度效應，而且幅度與實驗室中美國大學生受試者所得的幾乎相等。

許多我們將在第十章討論的機率推斷原理都源於實驗室研究，但同時也在田野研究中得到證實。例如，研究者探討過外科醫師、股票經紀人、陪審員、經濟學家和賭徒等人士，如何在其特殊的工作情境中做出機率推斷（Belsky & Gilovich, 1999; Gilovich, Griffin, & Kahneman, 2002; Kahneman & Tversky, 2000; Wagenaar, 1988）。行爲決策理論的原理，已應用在許多現實生活的情境中，例如，丹佛

市警察局用它來決定使用哪種型號的子彈最佳（Hammond, Harvey, & Hastie, 1992）；它也用來決定是否需要在亞歷桑那州的中部建造一座大壩（Hammond et al., 1992）。

實驗與非實驗結果的證據聚集，也成爲教育心理學領域的突出特點。例如，針對不同課程安排的實驗研究和田野研究都表明，早期語音教學有助於閱讀技巧的習得（Ehri, Nunes, Stahl, & Willows, 2001; Pressley, 2002; Rayner, Foorman, Perfetti, Pesetsky, & Seidenberg, 2001, 2002; Snow, Burns, & Griffin, 1998; Stanovich, 2000）。

簡言之，當前的心理學研究使用了不同種類的實驗技術和情境。儘管對許多問題的研究有時過分側重某類技術，然而與幾年前相比，心理學中不同研究方法的頻率分佈已經更加平均了。

◎向更精確的方法前進

對某一問題的研究，我們通常是從相對弱的方法向可以做出較強結論的技術變遷。例如，我們常受到某個非比尋常的有趣個案之激發，而對某一特定假設產生研究的興趣。正如第四章討論的，這反映出個案研究的眞正作用所在：爲進一步採用強有力方法來研究一個問題提供一些假設，同時激發科學家用更嚴格的方法做研究。緊接在個案研究之後，研究者多採用相關研究，以確認變數間是否眞正存在關聯，而不只是幾個特殊個案的巧合現象。如果相關研究證實變數間的關聯，研究者就開始嘗試採用實驗法操弄相關的變數，藉以找到變數間可能存在的因果關係。於是就產生一個從

個案研究到相關研究、再到實驗操弄變數的研究方法序列。儘管並非每個研究領域都遵循這一逐漸向精確研究方法前進的規律（有時不同類型的研究會同時進行），但這一研究方法序列的確比較常見。

在討論向更精確的研究方法進軍之前，我們必須糾正一些讀者可能從第五章的討論中產生的誤解，認為相關研究在科學中無用。的確，面臨要驗證一個因果假設，並且相關法和操弄變數的實驗法同時可行時，操弄變數法可能更受青睞；但是，這並不意味著相關研究對人類知識毫無貢獻。首先，許多科學假設都是以相關或缺乏相關的形式表達的，因此相關研究是直接在驗證這些假設。第二，儘管相關無法確定因果關係，但因果關係卻確保有相關。也就是說，儘管相關研究無法清楚地證實一個因果假設，但它可以起到排除因果假設的作用。第三，相關研究比表面上看起來更有用，這是因爲新近發展的一些複雜的相關設計，可以讓研究者做出有限的因果推論。我們在第五章中討論了偏相關這種複雜的相關技術，透過這一技術，我們有可能檢驗出是不是一個特定的第三變數造成觀察到的某個關係。

我們常用相關研究的最主要原因在於，有時出於道德的考慮，無法對一些變數進行操弄（例如營養不良或肢體殘障）。其他變數，諸如出生順序、性別、年齡等，則因無法操弄，而存在著本有的相關，涉及它們的科學知識，也因此必須建立在相關證據基礎上。當然，這一情況並非心理學領域獨有。天文學家顯然無法操弄所有影響其研究物體的變數，然而他們依然能夠從中做出結論。

有個實例可以闡明，在健康心理學中這種研究方法的演化。這個例子就是有關A型行爲模式與

心臟病之間的關聯（Curtis & O' Keefe, 2002; Friedman & Booth-Kewley, 1987; Friedman & Ulmer, 1984; Miller, Turner, Tindale, Posavac, & Dugoni, 1991; Wright, 1988）。最初，A型行爲模式這一概念源於兩位心臟病專家的觀察，這兩位醫師從一些病患之行爲發現一個穩定的模式，包括時間緊迫感、目標不定的敵意，以及對成就的極度追求。於是，一些醫師透過對少數個案的觀察，提出A型人格這一想法。這些個案研究催生了這個概念，但遠不足以作爲支持一種假設——一個特定的行爲模式是造成心臟病的原因之一——的有力證據。要證明這一點，我們需要的不僅是少數幾個個案研究，還需要由心臟病專家、生物化學專家和心理學家組成的研究隊伍數十年的努力。

對該問題的研究，迅速從簡單的病例累積向使用更精確之研究方法的方向不斷前進，因爲如果只是用個案研究法，這一假設的眞僞永遠難辨。研究者在朝向用一個比較精確的研究方法前進時，必須進一步發展出A型行爲的操作性定義。透過大範圍流行病學研究，研究者先在A型行爲和心臟病之間找到相關。接著，這種相關研究工作變得更爲深奧複雜。研究者利用複雜相關技術尋找潛在的第三變數。由於行爲模式與其他傳統心臟病風險因素（例如吸煙、肥胖和血液中的膽固醇水準）都有相關，所以很有可能A型行爲與心臟病的關係是虛假的。然而，研究結果表明，A型行爲可以是一個用來預測心臟病的獨立因素。因爲當其他可能影響該變數的因素以統計學的技術排除之後，A型行爲與心臟病之間的關聯仍然存在。

最後，研究者利用實驗研究操弄變數，藉以探討兩者之間是否可以找到因果關係。有些研究試

圖驗證，是不是某些生理機制影響了兩者的關係，並採用動物實驗——有些人認爲是「不是眞實生活」的研究方法。另一個實驗研究用罹患過心臟病的人做受試者。這些受試者隨機分派到兩個實驗組中，一組接受輔導，目的是幫助他們戒煙、過度攝取高脂肪食物等傳統心臟病的高危險行爲；另一組也接受輔導，不過目的是透過一種程序幫助他們減少A型行爲。三年後，發現後一組病患心臟病復發率明顯較低。簡言之，一系列證據聚集地指向支持A型行爲可作爲心臟病的重要原因之一這個假設。對這個問題的探研過程，可以說是一個很好的例子，用來說明研究如何逐漸地從有趣的個案研究轉化爲相關技術，再到更複雜的相關技術，最終達到可操弄變數的實驗研究。

我們從這個例子學到的最後一個教訓是，科學概念總是在不斷地發展之中。這個結論我們在第三章討論操作性定義時首次提出。最近的研究又傾向於表明，從**整體上**看，我們似乎過分簡化了A型行爲與心臟病之間的關聯。原因在於，只有該概念中的特定成分（特別是對抗性敵意）才與心臟病有關聯（Curtis & O'Keefe, 2002; Dembroski & Costa, 1988）。這個例子告訴我們，科學在前進過程中，如何發掘不斷地趨向具體的、精確的關係，以及理論概念如何細化。

對絶望的勸喻

證據聚集原則的最後一個意義在於，當某個問題的最初研究與結果之間有點矛盾時，我們不應

陷入絕望中。科學證據的累積過程，就像投影機慢慢地將一張內容未知的幻燈片調好焦距：起初，銀幕上的模糊影像可能代表任何東西；然後，隨著幻燈片一點點地調準焦距，雖然圖像仍不完全清楚，但許多原本以爲可能是什麼東西的假設已經否決掉了；最後，當焦距調準，我們就可以非常有信心地做出最終的鑒別。累積證據的過程在早期就如同調焦過程之始，幻燈片的模糊影像就如同互相矛盾的資料，或者可類比爲支持許多可能不同假設的證據。

所以，我們不應該對研究初期所得的相互矛盾之資料絕望，認爲永遠也無法找到眞相了。這種情形並非只在心理學才有，它也發生在許多比較成熟的學科中。實際上，大衆不了解的一點是，在任何科學中得到相互矛盾的資料是家常便飯，這些矛盾資料，充其量只不過是說明我們對那些資料缺乏充分的理解，它們可能源於隨機事件（第十一章將對此展開討論），或者源於實驗方法上的細微差異。

許多其他學科在取得內部共識之前，都經歷過一段不確定的混亂期。格蘭德維爾（M. Gladwell, 1996）敘述了有關對腦創傷患者進行適當緊急救治的近期思路發展過程。他講述了一個幸運的紐約病患，得到世界上最有名的專家加哲（Jam Ghajar）醫師爲他治療，加哲醫師一直致力於改變這個領域的臨床醫師持有的一個錯誤想法。格蘭德維爾描繪說，若干年前，當加哲和其他五位研究者在對一些創傷治療中心進行一項調查研究時發現，儘管類固醇已經反覆地證明無助於減少顱內壓力（而且會帶來潛在危害），仍有75%昏迷狀態中的病患是用類固醇來治療的。他寫道：

> 部分問題在於，在神經外科學這一領域中，難以對治療程序和方法達成一個可信的、科學的結論……再者，大腦的複雜性和神祕性，導致形成一個獎勵直覺行事的文化，並讓每個神經外科醫師堅信自己的經驗與其他人的經驗同樣可靠。（p. 39）

幾年前，講到自己同事的觀點停滯不前時，加哲寫到：「並不是神經外科專家很懶，而是相關的資料及資訊太多，以致造成混淆。」(p. 39)

簡言之，就如同心理學的許多領域一樣，這一領域也是到處都有人在做研究，但是焦距還沒有調好，也還沒想好怎麼看待這些資料，以便從中找到聚集的趨勢。於是在一九九四年，加哲和一些同事參加一系列的會議，試圖綜合當時所有的證據，藉以探看是否可以從這些證據中找到一個集中點。在這個腦創傷基金會（Brain Trauma Foundation）主辦的會議上，研究者就腦受傷處理的十四個方面考察了四千多篇科學論文。腦創傷基金會的執行主席描繪外科醫師如何展開工作：「他們所做的工作就是對科學文獻中記載的證據進行論證，一旦有人說『這是建立在我自身的經驗基礎上』時，每個人都嚷道：『不行，這不算數，等於沒說。我們想知道的是證據到底在哪裡。』」(Gladwell, 1996, p. 40）這種論證方式最終證明是有成效的：

> 參加會議的人發現，文獻中的確可以找到聚集的趨勢，因此出版了一本書展示所有的科學證據，以及腦創傷處理各階段的最先進治療法。這本指導手冊，是神經外科界首次成

功地建立的一套標準治療規範，如果美國大多數創傷中心都能嚴格按照這些規範做，可能每年可以多挽救上萬個病患。（Gladwell, 1996, p. 40）

儘管由加哲倡導的指導手册已經用於救死扶傷，有趣的是，當加哲本人談論這段醫學史時，他卻特別強調證據聚集的價值以及科學知識的公開性（第一章中討論過這一原則）：「人們總是想把它歸功於具體的個人……我猜想這就是人類本性吧。他們想說：『這是加哲的規範。他是個了不起的醫師。』但事實並非如此。這些規範是建立在當前最好的科學證據之上的。這是每個人都可以拿來用的規範。」（Gladwell, 1996, p. 40）

從這個例子我們可以清楚地看到，並非只有心理學研究結果表現出分散各地的現象，其他學科裡，也充滿因爲資料顯示的趨勢太過「模糊」以致難於作出結論的類似現象。然而，我們也清楚地看到，我們仍然可以找到一些方法，幫助我們找出一些初步的、暫時的結論。

正如第七章所討論的，從實驗結果中發現造成矛盾的原因，必將大大促進我們對某個現象的了解。心理學中，多數的幻燈片都還十分模糊，因此堅忍不拔就顯得加倍重要，因爲當那些困擾了我們幾個世紀的、有關人類行爲的複雜又有趣的問題最終得以解決時，我們也會得到雙倍的快樂。那些無法在模糊的氛圍裡工作的人，通常都不會從事科學工作，因爲科學就是充滿不確定性，它所呈現的模糊幻燈片可能破壞了某些人的美感。然而，科學家應有的態度是，對那些目前還很模糊但最

終將會變爲清晰的影像，保留一種強烈的好奇心。這就是布朗諾斯基（1973）所描繪的那種「科學是對我們可能知道的東西的一種奉獻，雖然在這個過程中，我們有可能犯錯。」（p. 374）

摘要

在本章中我們看到，爲什麼科學進步的突破模型不適合於心理學，以及爲什麼逐漸整合模型能爲心理學提供一個比較好的下定論的框架。證據聚集原則描述心理學是如何整合研究結果的：沒有任何一個實驗可以一錘定音，但每個實驗又幫助我們排除至少若干個其他的可能解釋，因而有助於追求事情的眞相。使用一系列不同的方法，可讓心理學家確信其結論是建立在堅實的實徵基礎之上的。最後，當研究概念發生變化時，它必須遵循關聯原則：新的理論不僅必須能夠解釋新的科學資料，還必須能夠對先前存在的資料提供一個解釋。

第九章　別相信「魔術子彈」

——多樣原因的議題

在第八章，我們探討了操作聚集的重要性，以及逐步朝向使用更精確的研究方法前進的必要性。這兩者對於在變數之間建立更精確的、單一的關聯非常有用。但是在這一章，我們將不再只看兩個變數之間的單一關聯，而把焦點放在思考另一個重點：人的行爲取決於多種因素。

任何具體的行爲，都不是單一變數引起的，而是由許多不同的因素決定的（Wachs, 2000）。當我們得出變數A和行爲B之間有因果關係時，並不意味著變數A就是引起行爲B的**唯一**因素。例如，有研究者發現，收看電視和學業成績之間有相關，但是他們不會就此認爲，看電視是影響學業成績的**唯一**因素。道理很簡單，學業成績在一定程度上，還受到大量其他因素的影響（例如家庭環境、學校教育的品質等等）。實際上，相對於這些變數，看電視只是影響學業成績的一個次要因素。同樣地，收看大量的暴力電視片，也不是使兒童表現出攻擊行爲的唯一原因，它只是衆多影響因素中的

一個。

和本書中談到的許多其他原則一樣，認識到事物具有多樣的成因這一點非常重要。一方面，它提醒我們不要過於依賴單一的原因解釋。因爲世界如此複雜，影響行爲的因素也相當繁雜。我們雖然可以證明某一變數引起某一行爲，並不代表我們發現了影響該行爲的唯一、甚至是最重要的因素。爲了充分解釋某種具體行爲，研究者必須探討各種不同變數的影響，並把這些研究結果整合起來，才能完整地描繪出所有與該行爲有關的因果關係。

另一方面，我們雖然說一個變數只是影響某一行爲的衆多因素之一，因此只能在很小程度上解釋這一行爲；但並不等於說這個變數不重要。首先，兩者的關係可能具有深遠的理論意義。其次，這一相關關係可能具有應用價值，尤其當那個影響變數是可以人爲控制時，如前面提到的電視暴力一例，我們是可以在電視上播放少一點暴力影片的。如果控制了這一個變數，能夠使每年的暴力事件降低1%的比率，我想沒有人會認爲它無關緊要。總之，如果這些負面行爲會對社會造成很大的影響，那麼懂得如何控制其中的一個成因，哪怕只是很小的一個，也相當有用了。羅森塔爾（R. Rosenthal, 1990）舉過一個治療心臟病的例子，在一個實驗中，某種治療方案能將患者存活率提高不到1%；然而即使這樣，這個結果也被認爲意義重大，以至於出於職業道德的考慮，實驗者不得不提早結束該實驗。原因很簡單：既然實驗治療結果這麼有效，對那些隨機分派在控制組的病患，讓他們仍然使用安慰劑，就顯得很不道德了。

交互作用的概念

原因多樣化的觀點引出另一個重要概念：交互作用（interaction）。這個概念在許多方法論的書上都會詳細的討論，所以這裡不再贅述，只是提一提：當影響行爲的一個因素和另一個因素共同起作用時，它們對該行爲會產生與各自單獨起作用時截然不同的影響。這就是我們常說的**交互作用**概念：一個變數的影響效果依賴於另一個變數的水準。賽蒙斯、博格森、卡爾頓—福特和布萊斯（R. Simmons, R. Burgeson, S. Carlton-Ford, and D. Blyth, 1987）指導的一個實驗提供這樣的例子。研究者考察一組青少年的學業平均成績，想看看生活事件（如轉學、青春期發育、早戀行爲、搬家和家庭破裂等）是否會對學業產生影響。他們發現上述生活事件加在一起，是影響學業成績的關鍵變數。

另一個例子是努特（Michael Rutter, 1979）對與兒童精神疾病相關因素的研究，他歸納道：

> 最初引人注目的發現是，在實驗中，那些單獨分離出來的長期壓力來源，各自並沒有誘發精神疾病的危險……這些高危險因素，當單獨看它們的作用時，都沒有與兒童的精神疾病連結在一起；這些兒童得精神疾病的風險，也不會比沒有家庭壓力的兒童高。然而，當任何兩種不同來源的壓力同時作用時，得病的風險就上升到原來的四倍。若是三種或四種壓力來源同時作用，得病的風險就又增大了好幾倍。很明顯地，這些長期壓力來源的共

同作用遠遠超過各自效果的累加。那是因為幾種併發壓力之間有交互作用，致使總體效應遠遠大於個別壓力效應的總和。（p. 295）

要理解交互作用如何產生的邏輯，我們可以像努特描述的，先想像有一把測量風險的量尺，尺上的數字80～110代表低風險，110～125代表中等風險，而125～150則代表高風險。假設我們發現兒童在無壓力情況下的平均風險指數爲82，在壓力因素A作用下的平均風險指數爲84，而在壓力因素B作用下的平均風險指數爲86。那麼當我們要同時研究因素A和因素B兩者對兒童的共同影響時，如果發現風險指數達到126，也就是說，併發的風險指數遠遠超過獨立研究單一因素時的結果，就說明了因素A和B之間存在著交互作用。

發展心理學中也有很多類似努特所述的例子。布瑞特米亞和雷米（Bonnie Breitmeyer and Craig Ramey, 1986）研究了兩組嬰兒，一組是非最佳預產狀況的嬰兒，另一組是正常嬰兒。在這兩類嬰兒出生後，再隨機分成實驗組及控制組兩組，然後對實驗組進行爲了防止出現輕微智力遲緩而設計的特別育嬰方案，控制組則沒有得到任何特殊照顧。當這些孩子長到四歲時，對他們的認知發展能力進行測試，發現在特別育嬰方案下接受照料的兒童中，非最佳預產狀況出生的兒童，與正常兒童在認知能力上沒有明顯差異。但是，沒有得到特殊照顧的控制組中，非最佳預產狀況出生兒童的表現，低於正常兒童的認知發展水準。該研究中，生理和環境因素的交互作用說明，一個複雜的行爲

結果（認知發展）是由多樣因素決定的。非最佳預產狀況出生的兒童又得不到適當的照顧時，就會出現最糟糕的認知發展結果。研究者總結道：「這個研究結果支持這樣的一個理論架構，即對於那些在社會經濟條件低下的家庭中成長的兒童而言，一出生時的生理缺陷會和後天不良的環境因素產生交互作用，從而成為他們認知發展的一個累積性危害因素。」（p. 1151）

同樣地，另一項有關驗證憂鬱症的體質—壓力理論研究（Metalsky & Joiner, 1992）表明，消極的生活事件與三種脆弱的心理因素相結合，會導致非常高的憂鬱症發病率。這三種因素是歸因方式（傾向於把消極事件歸因為一些穩定的、整體的因素）、對自我進行消極推斷的總體傾向，和對任何行為後果進行消極預測的總體傾向（參見 Alloy, Abramson, & Francis, 1999）。

很多消極行為和認知後果，都是由相似的邏輯所造成。例如，兒童攻擊行為是遺傳基因和不良社會環境相互作用造成的（Pennington & Ozonoff, 1996）。同樣地，佩迪特、貝茲、道奇和米斯（G. S. Pettit, J. E. Bates, K. A. Dodge, and D. W. Meece, 1999）發現，青少年成長的早期，如果在不受管教的情況（例如下課後或不在家中受父母監管時）下，有大部分時間是與同儕交往，**同時**在家裡也很少受到父母之監管的話，更易於把自己的問題表達出來。

正面的結果也可用多樣因素及其之間的交互作用進行解釋。耐特、強森、卡羅和艾森博格（G. P. Knight, L. Johnson, G. Carlo, and N. Eisenberg, 1994）在研究六至九歲兒童的利他行為時，探討了影響兒童助人行為傾向（如捐款給有需要的兒童）的心理因素。他們發現一些變數，諸如同情心、情意推理

和對金錢的認識等，當單獨看待它們的作用時並不起眼，它們和利他行爲之間的相關很低。然而一旦將這些變數放在一起看，對利他行爲就有很高的預測能力了。例如，具有較強的同情心、情意推理，並對金錢有所認識的兒童，捐款的可能性是在這些變數上表現較低之兒童的四倍。

因此，原因多樣化這個概念可能比你最初想到的複雜得多。我們不僅要追蹤並測量出影響問題行爲的種種可能因素，還必須研究這些變數放在一起會產生什麼作用。

單一因果解釋的誘惑

複雜事件是由多樣的原因決定的，這個觀點乍看似乎很容易理解。實際上，當問題沒有什麼太大爭議時，這個觀點**確實**很容易掌握和運用；但是，科學工作者的強敵——事先存在的偏見（參見第三章的討論），卻往往使我們忽視了原因多樣性這一原則。我們經常聽到人們在激烈地爭論一些話題，如犯罪的原因、財富的分配、歧視婦女和未成年人、貧困的原因、體罰的作用，以及納稅標準等時，大家的論點似乎總是讓人覺得這些問題是非常簡單的，而且導致結果的原因只有一個。這些例子進一步反映了尼斯貝特和羅斯（1980）的說法：「雖然人們通常也承認原因的多樣性，但顯然人們行動的表現往往和這一觀念背道而馳，他們傾向於用單一的因果關係解釋任何事物。在某種意義上說，人們的行爲好像是認爲各種原因存在著一種『水壓式』的關係，或者各種可能原因之間像

零和遊戲那樣，是一個相互競爭的關係的結果。」(p. 128)

零和遊戲——一個人贏的正是另一人輸的——常常是我們在討論那些容易激動的問題時採用的思維方法。在情緒高漲的影響下，我們通常會忘掉原因多樣性這個原則。讓我們來看看兩個敵對政黨是如何討論社會犯罪問題的，主張自由開放的人認爲，那些社經地位低下的人之所以會犯罪，是因爲他們本身就是惡劣社會環境（如失業、破陋的住所、缺乏教育和對未來喪失希望等）的受害者。而比較保守的人會爭辯說，大部分的窮人也都沒有犯罪啊！所以社會經濟條件並不是主要原因。因此他們認爲，個人的價值觀和人格特徵，才是決定犯罪行爲的眞正原因。在這場爭論中，雙方都沒有談到，造成犯罪行爲的原因既有個體的，也有環境的。

政治評論家科恩（Richard Cohen）也曾在文章中談到，我們如何根據先驗偏見做出「單一原因」的解釋，他引用愛荷華州一位六十三歲農場主人的例子。這位農場主在該州農業經濟進入蕭條期時，因爲背負沈重的債務，即將被銀行沒收農場。在感到還債無門的絕望中，他開槍打死借錢的那家銀行的經理，隨後在槍殺妻子之後自殺了。在討論這一事件時，無論鄰居還是傳統傳媒的看法都認爲，這個農場主是因爲遭受巨大的財務困難而「崩潰」。媒體報導寄予極大同情。科恩寫道：「那位農場主被描繪成一位勤勞的企業家，頑強地與大自然、銀行及芝加哥農產品交易市場的拍賣員那樣的人搏鬥。他誠實、節儉，是大多數美國人優點的化身——他是不靠他人的個體戶、自食其力，而且崇敬上帝。」(1985, p. 11)

但科恩質疑的是，如果這個人眞的只是因爲財務困難就殺人和自殺，那麼我們可否用同樣的單一原因理論解釋發生在貧民窟的殺人事件？「如果對於受困的農場主而言，這是其崩潰的原因，爲什麼發生在貧民區的情況就不同？爲什麼當我們提出貧窮、惡劣環境、三流學校和野蠻行徑是造成犯罪的原因時，會遭到如此多的責難？」（p. 11）科恩進一步指出，我們在做單一性原因解釋時產生的一個荒謬現象：我們是利用單一解釋鞏固既有的偏見。他認爲如果我們從一開始就認識到，發生在農場及貧民窟的殺人事件，都可能是多樣因素決定的，或許就可以避免犯下這一荒謬的錯誤。不論是愛荷華州的農場主還是貧民窟的人，他的殺人行爲都是受個人心理—生理特性和環境壓力共同影響的。沒有一個單一的原因可以單獨解釋犯罪事件，犯罪行爲是受許多因素影響的，包括環境因素，也包括個性因素。

在討論複雜的經濟形勢時也存在類似的例子。一個引起長達幾十年爭論、極具社會政策意義的重要話題是：爲什麼美國的貧富差距不斷地擴大（Beatty, 1996; Bronfenbrenner, McClelland, Wethington, Moen, & Ceci, 1996; Frank, 1999; Frank & Cook, 1995）。正如本書第六章「聰明漢斯」的例子，這一事實不是爭論的焦點，對這一事實的解釋才是。從一九七九年至今，美國全體男性的眞正收入（即排除通貨膨脹之後）下降幅度超過10%（Cassidy, 1995; Mishel, 1995）——儘管也有個別小群體做得很出色，收入有所增加。自一九七九年以來，持有研究生文憑的工作者群體的眞正收入增長了8.8%，但是「大專教育程度」的大多數工作者的眞正收入卻下降了15.9%（Beatty, 1996）。在美國家庭中占20%的中產

階級家庭的平均收入（同樣是排除了貨膨脹之後），從一九七三到九三年間大體上保持一致，主要是因爲這段時期大量家庭婦女湧入勞動力市場，令雙薪家庭增加了。婦女融入勞動力市場這一趨勢，其實主要是因爲受到工資水準停滯不前或不斷下降，而不是納稅增加的影響，靠丈夫一個人的單薪已經無法維持生計所造成的。因爲根據收入而定的稅率，從一九七九年（31.1%）到九三年（30.9%），一直保持在非常穩定的水準。

與中產階級和貧困階級的收入之停滯不前甚至不斷下降相比，恰恰是在一九七三到九五年這一時期裡，許多有錢人的財富暴漲（Frank, 1999）。人口總數裡收入最高的1%，從七七到九〇年的十餘年間，收入增長了74%（排除通貨膨脹之後的資料，Slemrod & Bakija, 1996）；與此同時，美國中產階級的收入只增長3%；而在人口總數中占20%的處於社會底層的人，收入卻下降了13%。一九七七年，社會上最富的20%賺的錢是最窮的20%的四倍。到九一年，此一比例增加到七倍（Frank & Cook, 1995）。一九七三年，最富的20%收入占美國國民收入總額的43.6%，到九三年，該比例已經達到48.2%（Cassidy, 1995）。

雖然這些百分比的差異看似不大，卻帶來深遠的社會影響。正如卡西迪（J. Cassidy, 1995）指出的：它意味著「在每年的收入中，由中產階級流入富人階級之口袋的錢大約有兩千七百五十億美元——平均每個中產階級的家庭要爲這一趨勢付出四千五百美元。」（p. 118）這已足夠讓大多數美國人的生活產生真正的變化。舉個例子，這將導致大部分公立大學的學費收入大大減少，因爲有經濟

能力供應子女上大學的家庭變得比較少了。

財富從市民中的一個階級大規模地轉移到另一階級，這個現象可能產生的社會影響，引發了一場極富爭議的、有關背後原因的政治辯論。爭辯最引人注目之處是，論者都只關注單一的原因，每一方只以某個原因爲立論基礎，然後千方百計地攻擊所有認爲其他因素是原因的觀點。事實上，計量經濟學研究（Beatty, 1996; Cassidy, 1995; Frank & Cook, 1995; Mishel, 1995）已經重點探研了四個變數（研究者提出的變數的數目遠不只四個，但這四個得到最廣泛關注及研究）。其中一個因素是科技。例如，論據之一是，懂得使用電腦的人提高了生產效率，工資也因而增加；與之相應的是，電腦替代掉很多沒有特殊技能的工人（如郵件分類員、銀行出納員等等），因而降低了他們的工資收入（Cassidy, 1995）。爭論的第二個因素是新移民的不斷湧入美國，這些多是缺乏技能的工人，他們使得已經很低的工資水準繼續下滑，因爲沒有技能的工作本就供過於求了。第三個原因是全球化，它助長了進一步的收入不均，因爲公司可以搬到國外，只需付很低的工資，就可以雇用一些沒有技術及低技能（但會逐漸地變成有技能）的工人，更加重本國低技能勞動力的過剩（Cassidy, 1995）。第四個原因是工會的影響力不斷下降，而大企業的控制力卻逐步增長。論據就是，在一九九〇年代，勞工的罷工事例減少，資方的罷工事例（即企業放棄某一區域爲生產基地，因爲在其他區域生產可以取得更高的資本報酬率）卻增加，從而提升了資本的價值，降低了勞動力的價值。

經濟學研究這四個變數時到底發現了什麼？你猜對了。這四個因素共同造成我們社會正在崛起

的不平等。這個例子也闡明了先前提到的、交互作用的概念。卡西迪（1995）在文章中指出：「某些因素可能有交互作用並相互強化。隨著全球競爭的日益激烈，企業經理人將更傾向於不理工會的加薪要挾，將資本投入可以取代人的電腦技術。同樣地，公司遷移的威脅和外來廉價勞工的增多，都會導致工會力量的進一步削弱。」（p. 122）

實際上，和經濟學的問題一樣，心理學研究的複雜問題也都受到多樣因素的影響。以學習障礙爲例，這個問題教育心理學家、認知心理學家和發展心理學家已經廣泛地研究過。結果發現，腦部的病變與學習障礙有關（Galaburda, 1994; Hynd, Clinton, & Hiemenz, 1999）。也有研究發現，學習障礙有一部分與遺傳有關（Cardon et al., 1994; Olson, 1999）。這兩個研究結果看起來好像可以讓我們做出一個結論：學習障礙僅是生物—腦方面的問題。如果我們很快就下這一結論，那就錯了（Sternberg & Spear-Swerling, 1999）。因爲進一步的研究發現，造成學習障礙的部分原因是，早期學校教育中缺少某些教導體驗（Pressley, 2002），以及家庭環境不好（Senechal & LeFevre, 2002; Snow, Burns, & Griffin, 1998）。學習障礙因此不是單一原因引起的；相反地，它是先天與環境因素交互作用的結果。

摘要

本章內容雖然簡單，卻非常重要。考察行爲的原因時，要依照多樣性原則思考。不要陷入類似

某一具體行爲只由某一特殊原因造成的這樣一個陷阱。大部分複雜的行爲都是由多樣因素決定的，各式各樣的因素共同起作用才會引發某一行爲。有時多個因素聯合在一起時會產生交互作用，也就是說，變數合在一起的整體效應，遠遠大於我們試圖只研究各自在單一變數因果研究中找到的單一效應之總和。

第十章　人類認知的罩門

——機率推理

問：「男人比女人高，對嗎？」

答：「**對**。」

問：「所有男人比所有女人都高，對嗎？」

答：「**不對**。」

答對了！信不信由你，在本章中，我們要花一些篇幅討論，從剛才的回答中看出你已經知道答案的問題。但是，先別因此就跳過這一章！因爲以下當我們要解釋一個看起來非常簡單的原則時，你會發現一些新東西。

回答第一個問題時，你給的是一個肯定的答案，這是因爲你沒有把「男人比女人高」這句話理

解成第二個句子所說的「所有的男人比所有的女人高」。你把第一句問話正確地理解爲「男人有比女人高的趨勢」，因爲每個人都知道，不是所有男人都比所有女人高。你理解到那句問話反映了一個**機率趨勢**（probabilistic trend），而不是任何情境都適用的事實。機率趨勢的意思是有較大的可能性，但並不是說在所有情況下必然如此。也就是說，性別和身高的關係要用可能性和機率的詞語，而不是用必然的字眼描述。在自然界中很多關係的本質也是機率性的，例如：接近赤道的地方天氣比較熱；每家的孩子數目不超過八個；地球上大部分地方的昆蟲數量比人多，這些都是可以用統計學概念表達的趨勢，但是當中的每一句話都不是絕對的，仍然可能有例外。因爲它們是機率的趨勢或定律，而不是在每一個別情況下都成立的關係。

事實上，心理科學揭示的所有事實和關係都是用機率闡述的。這一點，心理學不是唯一的學科，在其他學科裡，很多定律和關係也是用或然率、而不是用必然性闡述的。例如，人口遺傳學的所有子學科都是基於機率關係；物理學家告訴我們，原子中電子的放電分佈也是透過機率函數描述的。

的確，心理學揭示的大部分機率趨勢比其他學科弱一些。在心理學中，各種行爲關係都是以機率形式描繪的，然而這一事實並沒有使它與其他學科產生天壤之別。布朗諾斯基指出（1978a），隨著科學不斷地開拓出新的研究領域，愈來愈多的科學定律都將採用機率術語描述，對於這一點，許多人不能接受：

如果我說，一週中，經過前六天的好天氣之後，星期天總是會下雨，就會被認為是一個定律，因而受到重視。但如果我說，一週中，經過前六天的好天氣之後，星期天會下雨的可能性比不下雨來得大，就會是個不太令人滿意的說法，並且人們會理所當然的認為，我沒有真正找到一個定律。因為定律與通常人們習於從科學中得到那種決定性的「是」或「否」的結論是不謀而合的。甚至如果我說，一週中，經過前六天好天氣之後，十個星期天裡中有七個會下雨，你可能會把它當作統計數字來接受，但還是不能讓你滿意，因為它不是一個定律。它看上去就是不如定律那樣有威力。然而，這純粹是一種偏見……我在這裡解釋的機率理念你不難理解，但是它很新且不為人熟知。我們過去沒有接觸過……雖然生活在充滿有時和也許的世界裡，我們卻希望搬到和總是與一定住在一起。……然而我認為這個困難僅僅是一種習慣。一旦我們願意，一旦我們需要熟悉這些新的觀念，就會對它們逐漸熟悉。我們現在已到非熟悉這些機率理念不可的時刻了。（pp. 81-82, 94-95）

「張三李四」統計學

在這一章，我們將盡力使你更能把握「有時和也許的世界」，因爲了解心理學，必須熟悉本章討論的主題：機率推理。

眾所周知，醫學科學中很多結論是用機率趨勢、而不是必然性來描述的。吸煙會導致肺癌和種種其他健康問題，大量醫學證據都證明了這個事實。那麼每個吸煙的人都會得肺癌，而每個不吸煙的人就一定不會得肺癌嗎？大家都知道並不是這樣的。這種關係就是機率關係。吸煙在很大程度上增加患肺癌的機率，但並不必然導致肺癌。醫學科學可以很有信心地告訴我們，與不吸煙的人比較，吸煙者會有更多人死於肺癌。但是這個資訊不能告訴我們是哪些人會死，這種關係就是機率關係；它並不是在每個情況下都適用。我們都知道這一點，不是嗎？恐怕未必！我們經常看到下面這樣的場景：一個不吸煙的人引證吸煙導致肺癌的統計資料，企圖說服一個癮君子戒煙，得到的結果僅僅是對方的反唇相譏：「喂，走遠點！你看那個店鋪裡的老喬，他從十六歲開始每天抽三包駱駝牌烈煙！現在他已經八十一歲了，看上去還非常健壯！」任何人聽了這段話就只能得出一個結論：這個個案已經推翻了吸煙與肺癌的關係。

令人吃驚和沮喪的是，這種反駁手段屢試不爽。每當一個單一個案用來證明機率趨勢無效時，很多人常常點頭表示贊同，這反映出他們沒有正確理解統計定律的本質。如果人們認爲單單一個例子就可以使一個定律失效，他們一定認爲這個定律應該在任何情況下都適用。簡言之，他們錯誤理解了機率定律的性質，即使是最好的趨勢，也會有少數的「張三李四」與之背道而馳。以吸煙的例子來說，活到八十五歲的人中只有5％是吸煙者（University of California, Berkeley, 1991）。或者從另一方面來看，活到八十五歲的人中有95％屬於從不吸煙，或在一段時期內吸煙但最終停止者。連續從

未間斷地吸煙會明顯減少壽命（University of California, Berkeley, 1991），當然的確有少數長期吸煙者能活到八十五歲。

借用心理學家尼斯貝特和羅斯（1980）的術語，我們將把類似「老喬」的故事稱爲「張三李四」統計學原理的一個應用範例：由於一個人知道幾個「張三李四」與某個成熟的統計學趨勢相違，這個趨勢就會遭人懷疑。例如，我們經常聽到類似的話——「你說在服務業中就業機會在擴大；而在重工業中正在縮小？這不對，我知道某人上週四就在一個鋼鐵廠找到一份工作。」「你說與三十年前相比，家裡的孩子少了？你在胡說！隔壁的年輕夫婦已經有三個小孩，而他們還不到三十歲。」「你說通常孩子都會傾向與父母相同的宗教信仰？但據我所知，我的一個同事的孩子就在前幾天改信了另一個宗教。」

當我們面對和過去持有的觀念矛盾、又是強有力的統計證據時，無所不在的張三李四總是會立刻跳出來否定這些證據。因此可以這樣說，實際上人們比我們想像的知道得更多，他們只不過順手把張三李四當作工具，將違背他們觀念的事實否決掉。然而，研究人類決策和推理的心理學家的研究結果表明，人們之所以使用張三李四，不只因爲它在爭論時是一個很有用的戰略手段；這一錯誤的爭論模式之所以應用得如此頻繁，主要反而在於人類不知道如何處理機率性資訊。決策心理學的最新研究表明，機率推理可能正是人類認知的罩門。

機率推理與對心理學的誤解

在科學、技術和人文事物的許多領域中，都涉及機率性思維。所以，我們也沒有什麼特別的理由認爲這種思維對理解心理學比對理解其他學科更重要。然而，由於人們在運用機率性資訊時存在的問題，心理學的研究結果常常被誤解。我們都理解「男人比女人高」是一個機率趨勢的陳述，所以不會因爲一個例外（某個男人比某個女人矮），就認爲這一陳述是錯的。很多人也能以同樣的方式理解「吸煙會導致肺癌」的陳述，儘管對那些不想相信生活習慣會致他們於死地的癮君子來說，「老喬」可能還是具有說服力的。然而，與這些例子的陳述相似、不過是關於行爲趨勢的機率陳述，卻會引起很多人的懷疑，而且常常在一開始就被他們用一個張三李四個案打發掉。很多心理學教師在討論某些行爲之間關係的證據時，往往得到同一個反應。例如，教師可以呈現如下的事實，兒童的學業成就和家庭的社經地位及父母的教育水準相關。但這個事實常常會遭到至少一個同學的反對，他會說，他有個朋友是全國知名的學者，但他的父親不過中學畢業。甚至那些理解吸煙—肺癌例子的人，對這一問題也變得搖擺不定了。

人們從沒想過用張三李四的論據反駁醫學和物理上的發現，卻習慣用它們來駁斥心理學的研究結果。大多數人能理解醫學科學提出的治療、理論及事實是機率性的。例如，他們理解一種藥對一群病患來說，並不是每個都有療效，而且醫學也經常不能事先告訴我們對哪些病患有療效。通常可

以說的是，一百個病患接受A療法，一百個病患不接受任何治療，在一段時期之後，接受A療法的這一百個病患，**整體上**比沒接受A療法的一百個病情好一些。沒有人因爲這是一個機率陳述，因而不是在所有情況下都適用，就懷疑這一治療的價值。許多心理學的研究結果和心理治療的療效，也存在類似的情況。然而，一旦心理學研究結果和心理治療療效不能在所有情況下適用時，就常常引發人們對心理學的進展產生極大的失望和輕蔑。一旦面對心理學的議題，人們常常忘記一個最基本的原則：知識不需要百分之百確定後才是有用的——即便某些知識不能預測個體的具體情況，但如果能對一些群體的總趨勢有預測能力，也是非常有益的。基於群體的特徵所做的結果預測，常常稱爲**總體統計數字**（aggregate）或**精算式預測**（actuarial prediction，我們將在下一章詳細討論這一概念）。

由於這些原因，全面地理解機率推理對理解心理學非常關鍵。在這裡，我們也看到一個激起人們深思的諷刺現象，心理學很可能是人們不能作統計式思考的最大受害者，然而心理學家卻是對人類機率推理能力進行最多研究的人。

機率推理的心理學研究

過去二十年來，普林斯頓大學（Princeton University）的卡尼曼（二〇〇二年諾貝爾獎得主，參見 Mac-Coun, 2002）、密西根大學的尼斯貝特和已故的特維斯基（Amos Tversky）等心理學家的研究，徹底改

變了我們對人類推理能力的認識。他們在研究期間發現，很多人不具備某些機率推理的基本原理，或即使有也不夠用。正如學者經常指出的，這些基本原理在人們心裡沒有發展好是不足爲奇的。作爲數學的一個分支，統計學是最近才發展出來的（Hacking, 1975）。反而是運用機率的遊戲，要比機率定律早出現幾個世紀。這又是一個例子，說明個人的經驗不足以讓人們獲得對世界的基本理解（參見第七章）。成千上萬的賭徒及其「個人經驗」，並不足以讓我們發現機率遊戲的本質；反而是透過對機率定律的正規研究，才讓我們找到人們一直在玩的機率遊戲背後的運作方式。

社會變得愈複雜，我們就愈需要機率性思維。如果一個普通市民想對其生活的社會有基本的理解，那麼，他在最低的限度上，必須具備可以運用統計式思維的最基本能力。

「爲什麼他們要提高我的保險費？」「爲什麼張三的保費比李四高，是社會保險局搞錯了麼？」「我們這一州的樂透背後有黑幕？」「犯罪事件到底是在增加還是減少？」「爲什麼醫師要安排這些檢查？」「爲什麼在歐洲可以用一些珍奇的藥，而美國就不行？」「在相似的工作崗位上，女人賺的眞的比男人少嗎？」「國際貿易眞的減少美國人的就業機會，並降低他們的薪水嗎？」「日本的教育成就比美國好嗎？」「加拿大的衛生保健眞是比美國好又低廉嗎？」這些問題都問得很好——都是攸關我們社會及其如何運作的具體而實際的問題。要理解每個問題，我們就必須運用統計式思維。

限於篇幅，本書當然不能全面覆蓋統計式思維的所有討論。然而，我們將簡要地討論某些關於機率推理較爲普遍的陷阱。學習機率性思維技巧的最好方法是：理解並警覺人們使用統計推理時最

常犯的錯誤。況且，這些錯誤有許多可以幫助我們進一步看到心理學研究結果及理論的重要性。

◎未充分利用機率性資訊

在心理學領域中，有個發現已經得到多次重複驗證：一個具體單一事件的資訊，往往可以壓倒比較抽象的機率性資訊（第四章中討論的鮮活性問題）。忽視機率性資訊的例子比比皆是，而且不局限於科學知識不豐富的外行人。卡斯塞爾斯、謝諾博格和格瑞博維斯（W. Casscells, A. Schoenberger, and T. Graboys, 1978）在四個哈佛醫學院的教學醫院中進行一項研究，他們問二十位醫學專業的學生、二十位內科主治醫師和二十位行政人員下面一系列問題：如果在每一千人中有一個人是愛滋病毒（HIV）的帶原者，再假設有一種檢查可以百分之百診斷出眞的帶原者；最後，假設這個檢查有5％的可能性誤認非帶原者爲帶原者。也就是說，這項檢查在非HIV帶原者中，會錯誤地檢測出有5％的人是帶原者。假設我們隨便找個人進行這項檢查，得到呈陽性反應，亦即此人爲HIV帶原者。假定我們不知道這個人的病史，那麼他眞的是HIV帶原者的機率是多少？

一般人會回答95％，正確的答案是接近2％。醫師對一個診斷爲陽性反應的病例到底是否爲確診之統計推斷往往過分高估，這是因爲他們一方面過分重視個案資訊，另一方面又過分忽視基本比率資訊，從而過分高估陽性反應測驗結果眞正代表的患病機率。少許的邏輯推理，可以幫助我們說明基本比率對機率性思維的重要作用。一千人中只有一個人會確診是HIV呈陽性反應。如果另外

九九九人（不患有這種病）也進行檢查，由於這一檢查有5%的虛報率，其中將有接近五十人（999×.05）檢查出爲帶原者。這樣一來，呈陽性反應的人就會是五十一個。由於這五十一人中只有一個人眞正是ＨＩＶ呈陽性反應者，此人確診得病的機率其實只接近2%。簡而言之，這一假想題的基本比率就是大多數人並非帶原者。這個事實和相當大比例的虛報率放在一起考慮就能算出，在絕對數量上，大部分呈陽性反應的人並非帶原者。

儘管參與卡斯塞爾斯等人研究的醫師，很快意識到以上機率邏輯的正確性，但他們最初的直覺反應卻是忽視基本比率，並過分看重臨床檢測的證據。簡單地說，事實上醫師早知答案不會是那麼高的百分比，卻從資料中得出錯誤的結論。心理學家把這類問題稱爲**認知錯覺**（cognitive illusions，參見 Kahneman & Frederick, 2002）。在認知錯覺中，即使人們知道正確答案，也會因爲問題的問法而作出錯誤的結論。

在這裡，我們提到的所有例子都是認知錯覺，因爲它們都落入人類推理的一個陷阱：過分看重個別事件提供的證據，而忽視統計式資訊。對大多數人來講，個案證據（實驗室的研究結果）好像是抓得住的、具體的，而機率證據則好像是抓不住、不確定的。當然，這種理解是錯誤的，因爲個案證據本身就是機率性的。一個臨床檢驗會在一定的**機率**內誤診某人有某種疾病。前面舉的例子就是有兩種機率同時作用——對個案證據做出正確或錯誤診斷的機率和過去經驗提供的先驗機率（prior probability），要做出正確的判斷必須將二者結合起來。結合這些機率的方法有的是正確的，有的是

錯誤的；特別是當個案證據給人一種它比較抓得住、比較具體的錯覺時（請回憶一下第四章討論的鮮活性問題）——人們往往就以錯誤的方式結合資訊。這個具體的機率推理失敗，說不定正是阻礙心理學知識應用的禍首，因爲心理學的知識，正是經常用機率的形式表述人類各種行爲之間的關係。

科普作家科爾（1998）要我們想像下面兩種情況。一種是用抽煙的死亡率勸人不要吸煙，這是一種最常見的勸人方式。第二種方法則更生動，讓吸煙者想像在每一八、二五〇包煙中有一包是與衆不同的——它裡面裝滿炸藥，當吸煙者打開時就會被炸死。我們絕對知道哪一個更具有效果——然而它們表達的卻是同一個事實。

最後，愛滋病的例子是體現機率性思維在美國社會之重要性的一個好例證。有些人可能認爲這是一個虛構的故事，但是，很多年來，在某些職業和政府部門中確實強制實施愛滋病檢查，而且這一問題已經引起激烈的爭論，可悲的是，這一強制檢查恐怕會繼續成爲社會問題。正如數學教授斯廷（Lynn Steen, 1990）辯稱的：

> 對強制性愛滋病檢查的持續爭論，提供一個很好的例子，說明了隱藏在這些爭論之下的有關數量的問題……檢查總會有數量很少的（或許2%）錯誤……因此大眾以此推斷這一檢查的正確率是98%。但是，因為在普通人中出現愛滋病例少於檢查中錯診的數量，所以對隨機取樣的市民進行任何大規模的檢查，被誤診為有愛滋病的可能性，其實遠遠大於

他真的是帶原者。（p. 218）

◎未充分利用樣本大小資訊

請大家想想下面兩個由特維斯基和卡尼曼（1974）提出的問題：

1. 一個小鎮裡有大小兩家醫院。在大醫院裡每天大約有45個嬰兒出生，小醫院約15個。正如大家知道的，大約有50%的嬰兒是男孩。當然，眞正的百分比每天都不一樣，有時候高於、有時候低於50%。在一年內，每家醫院都記錄了出生的男嬰比例高於60%的天數。你認爲哪一家醫院記錄的天數多？

a. 大醫院

b. 小醫院

c. 大約一樣

2. 假設一只缸裡裝滿球，其中三分之二是一種顏色，三分之一是另一種。一個人從缸裡拿出5顆球發現有4顆是紅色，1顆是白色。另一個人從裡面拿出20顆球，發現有12顆紅色，8顆白色。哪一個人比較確信缸裡有三分之二的球是紅色，三分之一是白色，而不是相反？這兩人對他們的猜測會有什麼樣的確信度？

針對第一個問題，大多數人回答「基本一樣」。不這麼回答的人，各有一半選擇大醫院和小醫院。但正確的答案是小醫院，所以接近75%的受試者都給出錯誤的答案。這是由於人們沒有認識到樣本大小在問題中的重要性。當其他的因素不變時，較大的樣本總是能夠更精確地估計出樣本母體的眞正數値。也就是說，在任何一個指定的日子，較大的醫院由於有較大的樣本，男嬰出生的機率傾向於接近50%。相反，小的樣本總是傾向距離樣本母體的眞正數値比較遠。因此，小醫院會有更多的天數記錄了與樣本母體的眞正數値相矛盾的男嬰出生比率（60%、40%、80%的男嬰，等等）。

回答第二題時，大多數人認爲5顆球的樣本提供了更令人信服的證據，可以證明缸裡的球大多數是紅色。事實上，利用機率性思維，恰恰得到相反的結果。在5顆球的樣本中，缸裡眞有三分之二爲紅球時，抓出4顆紅球、1顆白球的機率，與眞有三分之一爲紅球時的機率之比例是8：1。而在20顆球的樣本中，缸裡眞有三分之二爲時，抓到12顆紅球、8顆白球的機率，與眞有三分之一爲紅球時的機率之比例是16：1。儘管在5顆球的樣本中，抓出的紅球比例較高（80%），並不能抵消另一個取樣大小爲其四倍的樣本，進行推斷時有較大的可信度這一事實。然而，大部分受試者在這個題目作出判斷時，多因爲5顆球的樣本中紅球有較高的比例，而沒有充分考慮到20顆球的樣本有較大的可信度這一點。

認識到樣本大小對資訊可信度的作用，是對證據進行評估時必須注意的基本原則。這一基本原則，固然經常應用於許多不同的領域中，不過它對評估行爲科學的研究結果尤爲重要。不管我們是

否意識到，我們對大一點的群體，的確會持有一些泛化的想法。然而，我們最堅定的一些想法是建立在那麼脆弱的事實基礎上，倒眞是我們很少察覺到的。把對幾個鄰居和幾個同事的觀察，以及在電視新聞上看到的一些趣聞軼事放在一起，我們就覺得自己有資格對人性或「美國人」評頭論足。這裡，我們只是指所用樣本的大小，還完全沒有談及樣本是否具有代表性。那當然又是另一個問題了。

◎賭徒的謬誤

賭徒的謬誤是指人們認爲過去發生的和將要發生的兩個相互獨立的事物之間，存在著某種關聯的傾向。如果一件事發生之後，對另一件事發生的可能性沒有任何影響，這兩個事件就可說是相互獨立的。大多數裝置有隨機設備的賭博遊戲就有這個特點。例如，每一次在輪盤上出現的數字，並不依賴於上一次出現的數字。在輪盤上有一半的數字是紅色的，另一半是黑色的（爲了簡單起見，我們忽略綠色的零和雙零這兩個數字），這樣，旋轉後得到黑色和紅色數字的機率是一樣的，所以每一次旋轉後都有可能得到紅色的數字。然而連續五、六次是紅色數字後，很多賭徒都迅速的將籌碼轉押黑色，認爲下次開黑色數字的可能性較高。這就是賭徒的謬誤：事件之間原本相互獨立，卻認爲前一個事件的結果影響了後一個發生的機率。賭徒的這一想法是錯誤的。輪盤對先前發生的事情是沒有記憶的。即使一連串有15個紅色數字出現後，下一次開出紅色數字的機率仍是50%。

你可以用擲硬幣證明在大衆的心目中也存在類似的謬論。如果你問一群人，連續五次擲硬幣得到的都是人頭朝上，下一次得到人頭朝上的機率是多少。一部分人會說不可能再出現人頭朝上。這又是賭徒的謬誤。每次擲硬幣都是獨立的事件。連續五次人頭朝上後，硬幣仍有兩面，所以在擲硬幣時，每一面都有相等的機率人頭會朝上。

不僅僅是沒有經驗的賭徒或初學者才會犯賭徒的謬誤。過去研究發現，即使是每星期玩賭博遊戲超過二十個小時的資深賭徒，仍然會表現出賭徒的謬誤（Wagenaar, 1988）。而且，我們要認識到這一謬論不僅僅局限於賭博遊戲，它還存在於任何機率起著重要作用的領域。換句話說，它幾乎發生在**每一件事**上。嬰兒的基因構成就是一個例子。心理學家、醫師和婚姻顧問常常看到已有兩個女孩的夫婦，在計畫生第三胎說，因爲「我們想要個男孩，這回一定是個男孩」。這就是賭徒的謬誤，生了兩個女孩之後生男孩的機率（接近50%）和生第一個時完全一樣。生了兩個女孩之後**不會增加**第三個是男孩的機率。

賭徒的謬誤存在於任何一個有機率成分的地方，例如在體育運動和股票市場。一些心理學家（Gilovich, Vallone, & Tversky,1985）研究了籃球運動中對「連投連中」或「手熱得發燙」的迷信，這一迷信是指，相信某個球員能夠變得「手熱得發燙」，並且接下來會連續投中（「球傳給他，他現在熱得發燙」）。研究者證實在籃球運動員及球迷中，連投連中的想法非常強烈。例如，在一個問卷調查中，91%的籃球迷認爲剛投中兩球或三球的球員，與剛有兩次或三次投不中的球員相比，在下一次

投籃時會有較高的投中機率；84%的球迷認爲，應該把球傳給剛剛連續投中兩三球的球員。當請球迷估計，假設一個球員有50%的命中率，那麼在他投中一球後，再投中的百分比是多少？一次沒投中後，再投時投中的百分比又是多少？結果，球迷估計前者是61%，後者是42%。研究者調查了費城七六人隊（Philadelphia 76ers）的隊員，結果發現，大多數（但不是全部）球員對連投連中所持有的強烈信念與球迷幾乎一樣（參見 Gilovich et al., 1985）。

但爲什麼我們要在賭徒的謬誤的大標題下討論連投連中呢？**因爲根本沒有連投連中這回事！**基勒維奇等人（1985）研究了費城七六人和波士頓塞爾蒂克隊（Boston Celtics）一九八〇到八一年球季中投籃命中的統計資料。在這一球季期間，球員的投籃並沒有序列關聯性（sequential dependency）。現在讓我們用非技術語言看看這代表什麼意思。

賭徒的謬誤是把獨立的事件加以連結，也就是認爲毫無關係的事件之間存在著序列關聯性。在統計學上，連投連中是指連續投中兩三球後，投籃的命中率高於連續兩三次沒投中的假設。基勒維奇等人（1985）計算了這個機率，發現沒有任何證據支持這個假設，例如，爾文（Julius Erving，以「J博士」的綽號聞名，是費城七六人隊中籃最多的紀錄保持者）的資料表明他在連續三次投中後，投籃的命中率爲48%，而連續三次沒投中後投籃的命中率爲52%；在連續兩次投中後的命中率爲52%，而連續兩次沒中後爲51%；在一次投中後的命中率爲53%，在一次沒投中後爲51%。簡單地說，無論前幾次投籃的情況如何，J博士的命中率都是接近50%——根本沒有連投連中這回事。

其他球員的資料也非常類似。霍林斯（Lionel Hollins）連續兩次投中後的投籃命中率是46%，連續兩次沒中後是49%；投中一次後的命中率是46%，和一次沒投中後的命中率完全一樣。這說明，不管霍林斯前幾次的投籃結果如何，他投籃的命中率總是接近47%。波士頓塞爾蒂克隊的罰球資料也說明了同樣的情況。例如，柏德（Larry Bird）在罰中一球後下一次罰中的機率是88%，而一次罰不中後是91%。阿奇巴（Nate Archibald）在罰中一球後，下次的命中率是83%，而一次沒投中後是82%。由此可見，在罰球中也沒有連投連中。相信球員「手熱得發燙」的想法，確實是賭徒的謬誤的一個例子，也就是說，相信事實上獨立的、毫無關係的事件存在著關聯。

有趣的是，賭徒的謬誤看來又一次驗證我們在第六章討論「直覺式物理學」時的結論——僅憑經驗無法告訴人們世界的眞相。基勒維奇等人（1985）測驗了大學籃球隊員，他們在球場中（沒人防守）練習十五呎投籃，讓這些球員針對一百次投籃的命中率打賭。由於隊員一般在這個距離能夠投中的比例約爲50%，因此可以保證他們不一定能贏。打賭的規則是，投中贏的要比沒投中輸的多一些。然而，球員可以選擇下多些賭注（這樣贏的多輸的也多）或少下些（贏的少輸的也少）。顯然，如果球員能夠預測自己的成績，就會贏得比較多；也就是說，當他們認爲投中的機率高時，就會選擇加大賭注；認爲投中的機率低時，就會選擇減少賭注。實驗結果表明，和職業球員一樣，沒有發現「手熱得發燙」的現象：一次或多次投中後，再投中的機率並不比一次沒中後再投中時更高。然而，球員卻都**認爲**有手熱得發燙這回事。他們在投中一球後，對下一次投籃所下的賭注，要高於沒

投中後。結果證明，球員**根本不能預測自己的表現**：他們預測的結果並不比隨機預測來得好。

賭徒的謬誤源於對機率這一概念的許多錯誤觀念。其中的一個錯覺就是，如果一個過程眞正是隨機的，就不可能出現有連續重複同樣結果或模式的排列出現，哪怕是一個不起眼的隨機事件（例如擲六次硬幣）。人們習慣低估在一個隨機排列中，連續重複同樣結果（HHHH）或樣式（HHTTHHTT）的可能性。由於這個原因，在製作一組眞正的隨機排列時，卻常常適得其反地製作出一個很少出現連續重複同樣的結果或樣式的排列。這是因爲，人們在製作時往往錯誤地讓可能的結果儘量輪流出現，以爲這樣才稱得上是隨機抽樣，當然就打亂了眞正隨機排列中內含的這種重複結構（Lopes & Oden, 1987; Nickerson, 2002）。

那些聲稱自己有通靈能力的人，可以輕易地利用人們這一錯覺。讓我們看看下面這個在大學心理學課中有時會進行的示範。讓一位學生準備兩百個分別由1、2、3三個數字隨機抽取的排列。完成之後，不要讓作演示的老師看到。接下來，讓這個學生集中精力在第一個數字上，由老師猜這個數字是什麼。當老師報告他的猜測之後，這個學生再向全班同學及老師公佈正確的答案。以此類推，猜完這兩百個數字，並記錄下老師猜對的數目。在實驗開始之前，老師聲稱有通靈能力，可以在實驗過程中用「通靈術」得知別人的想法。通常，老師會先問班上學生，他猜測的成績會是如何——也就是「命中」的百分比多少，這樣，他的示範才能成爲支持他有通靈術的堅強實徵證據。這時，通常都會有個修過統計課程的學生回答，因爲純粹隨機的猜測能得到猜中33%的結果，所以要

讓別人相信他有通靈術，猜中的比例就一定要超過33％，大約至少40％。班上大部分同學都會認同這個觀點。實驗結束後，老師猜中的比例果眞超過40％。這個結果令很多同學感到驚訝。

學生從這一示範中得到一些什麼是隨機的教訓，並且了解到僞裝通靈能力是多麼容易呀！在這個例子中，老師只是利用人們在製造「隨機」數字時，經常不讓連續重複的數字出現，以致常換來換去這一行爲現象。在眞正的隨機排列中，已經出現三個2之後，再出現2的機率是多少？其實還是三分之一，與出現1或3的機率一樣大；但大多數人在隨機製作數字時並非如此。即使只是出現兩三個相同數字的重複，人們爲了得到一個典型的隨機排列，常常會刻意交替給出不同的數字。在這個例子中，老師只要在每一輪猜測前，不去挑選那個學生在前一輪中挑選的數字，而從另外兩個數字中選一個就可以了。例如，如果那個實驗中的學生的上一輪數字是2，那麼老師就會在下一輪的猜測中選1或3；如果學生在上一輪說的數字是3，老師接著就會猜1或2。這麼簡單的把戲，通常可以保證猜中的機率高於33％——高於三個數字隨機猜測的準確率，根本不需要動用什麼通靈能力。

◎再談統計與機率

以上提到的這些問題，其實只是阻礙正確理解心理學的統計推理缺陷中的一小部分。有興趣的讀者可以參閱下面介紹的幾本書，它們在這方面提供了比較完整、詳細的描述：基勒維奇、格里芬

（D. Griffin）和卡尼曼編的《訣竅和偏見》（*Heuristics and Biases: The Psychology of Intuitive Judgment*, 2002）。另外，還有皮爾蒂利—帕爾瑪瑞尼（Piattelli-Palmarini）的《不可避免的幻覺》（*Inevitable Illusions: How Mistakes of Reason Rule Our Minds*, 1994），拜倫（J. Baron）的《被誤導的判斷》（*Judgment Misguided: Intuition and Error in Public Decision Making*, 1998），韋爾奇（D. A. Welch）的《決策，決策》（*Decisions, Decisions: The Art of Effective Decision Making*, 2002），吉格任瑟的《精算過的冒險》（*Calculated Risks: How to Know When Numbers Deceive You*, 2002）。它們都對統計與機率作了通俗的介紹（對沒有受過任何數學訓練的初學者來說尤其有用）。哈斯戴和道斯（R. Hastie and R. M. Dawes）的《在一個不確定世界的理性選擇》（*Rational Choice in an Uncertain World*, 2001）和拜倫的《思考和決策》（*Thinking and Deciding*, 2000）兩本書中的介紹雖然有點專業化，但仍具有很高的可讀性。

本章討論的機率性思考具有巨大的實踐意義。由於沒有充分運用機率性思考能力，醫師選擇了效果不好的治療方法（Baron, 1998; Dawes, 2001），人們不能準確地評估周遭環境帶來的危險（Margolis, 1996），在訴訟程序中專家資訊被誤用（Foster & Huber, 1999; Lees-Haley, 1997），政府和私人企業將數以百萬計的資金用於不必要的專案（Arkes & Ayton, 1999; Dawes, 1988, pp. 23-24），動物不斷被捕殺以至瀕臨滅絕（Baron, 1998; Dawkins, 1998），不必要手術的實施（Dawes, 1988, pp. 73-75），損失慘重的財政誤判（Belsky & Gilovich, 1999; Kahneman & Tversky, 2000；Thaler, 1992; Zweig, 2001）。

當然，我們不可能以短短一章全面地討論統計推理。本章的目的，只是想強調統計在研究及理

解心理學上的重要性。不幸的是，當碰到統計資訊時，我們還找不到一個放諸四海皆準的規則。不像科學思維中的其他部分比較容易掌握，而統計學中，功能性的推理技術可能需要透過正規學習才能掌握。幸運的是，現在大多數綜合大學和社區學院，都提供了入門水準的統計學課程，而且通常不需要以大學程度的數學課爲先修條件。上這類課程之前，讀者可以先閱讀我剛才推薦的那些書。

很多科學家都眞誠地希望科學知識也可以讓一般大衆知道及理解。然而，有時一門學科必須先掌握某些技術才能理解，而與這些技術相關的資訊又只有在正規學習中才能獲得。因此，如果我們讓人們誤以爲對這門學科之深入理解，是一般外行人都做得到的，在學術上就是不負責任的做法。統計學和心理學就是屬於這一類的學科。心理學家班歐（Alan Boneau, 1990）調查了心理學教科書的作者，請他們列出學生在學習心理學時需要掌握的最重要術語和概念。其中頻率最高的一百個，接近40％是屬於統計學和方法論的範疇。不精通統計和機率的人，如今已不可能成爲一個好的心理學家（Friedrich, Buday, & Kerr, 2000）。

不可否認地，本書的目的之一，就是要使心理學的研究廣爲讀者接受。然而，在建構心理學的理論時所用的實徵方法和技術，和統計學已密不可分（這點和很多領域一樣，例如經濟學、社會學和遺傳學等），如果我們認爲一個沒有統計學知識的人可以完全理解心理學，那就大錯特錯了。因此，儘管這一章非常粗略地介紹統計式思考，但主要目的是在強調要充分理解心理學，另外一門學科的知識是絕對不可少的。

摘要

就像大多數的學科一樣，心理學研究得出的結論是機率式結論——其推論的結果是在大多數的情況下會發生，也是較為普遍的，但這些結論絕不是在任何情況都會出現的。也就是說，它們不是百分之百準確的（就像所有醫學的治療方案那樣，它們只是在一定機率的意義上有療效）。儘管如此，根據心理學研究及理論所做的預測仍然是有用的。

妨礙很多人理解心理學研究的原因在於，他們很難用機率的術語思考。在這一章，我們討論了幾個相當精彩的研究例子，表明機率推理如何令大多數人止步不前：當人們遇到具體的、具有鮮活性的證據時，他們就將機率性資訊置之腦後了。他們也沒有考慮到，與較小的樣本相比，較大的樣本估計的母體數值會比較精確。最後，人們也表現出賭徒的謬誤：把原本無關的事件看成有關聯。這個謬誤是源於下一章將要討論的、人們更常有的一個傾向：在決定因果關係時，未能認識到偶然機遇所起的作用。

第十一章　心理學中的偶然

在上一章，我們討論了機率趨勢、機率性思考和統計推理的重要性。本章將繼續這一討論，但是重點放在對隨機性（randomness）和偶然（chance）這兩個概念的理解困難上。我們將強調人們經常由於不能領會偶然這一概念如何徹底地融合在心理學理論中，以致誤解了研究對臨床實踐的貢獻。

試圖解釋偶然事件的傾向

人類的大腦已經演化到一個程度，能讓我們不懈地尋求世界上事物存在的各種樣式。對於身邊發生的事物，我們總想尋找它們之間的關係、對它們進行解釋並尋求這些事物背後的意義。這顯然是一個極具生存適應的特點，它使我們能預測世界將發生什麼事，並且最終有可能找到爲什麼會出現某一樣式的精確解釋。當然，這種對樣式的探求是科學家的工作，但並不局限於科學家，我們每

個人每一天都進行著這種探求活動。心理學家曾經研究人們這種探求事物結構的強烈傾向，並指出它是人類智力的特性，而且此一活動是人類在資訊加工和知識獲得方面，能取得許多令人震驚的偉大成就背後的眞正原因。

然而，人類認識過程的這種極具生存適應性的特點並非十全十美，有時還會給我們帶來一些障礙。例如，當環境中沒有什麼東西可以讓人們進行概念化時，還一味地尋求用概念來理解事物，就是一種不良的適應行爲。那麼，到底是什麼東西在人類認識的這一突出特點上製造麻煩？又是什麼東西給我們尋求結構這一過程帶來困擾，以致阻礙了我們對事物的理解？你猜對啦！是機率。更具體的說：是偶然和隨機性。

偶然和隨機性是我們周圍環境不可分割的一部分。偶然和隨機性的規律是支配生物演化和基因重組的機制，物理學也運用偶然的相關統計規律來解釋物質的基本結構。自然界發生的很多事情都是穩定的、系統的、可解釋的因素與偶然因素綜合作用的結果。回想前面談到的一個例子：吸煙導致肺癌。生物學上系統的、可解釋的因素，把吸煙和某一疾病之間建立關聯，但這並不表示所有吸煙者都會罹患肺癌，這種趨勢是機率性的。也許，最終我們能解釋爲什麼有些吸煙者不會罹患肺癌，但是在現階段，這種不穩定性必須歸因於決定一個人是否感染某一疾病的大量偶然因素。

在上面的例子中，當我們說一件事決定於偶然時，並不必然表示它是**不確定的**，而只是說它是當下**無法確定的**（Alcock, 1981, p. 155）。例如，說拋擲一枚硬幣是偶然事件，並不是因爲當我們透過測

量拋擲的角度、硬幣的金屬含量以及許多其他變數等，都不可能確定拋擲的結果，實際上，這些變數的確決定了拋擲硬幣的結果。但正是因爲在每一次拋擲時，我們沒有比較簡易快捷的方法測量這些變數，所以才視之爲一個偶然事件，一次拋擲的結果不是具有不確定性，只是在當下無法確定。

世界上的許多事件不能只運用穩定的、系統的因素完全解釋清楚，至少現在還不能，我們必須求助於在事件中也起作用的許多偶然因素。不幸的是，大多數人都覺得這樣做不能令人滿意。當然，如果這種質疑激勵了我們對一種現象做進一步的探究，這是一件好事。但是如果它促使我們匆匆忙忙地接受某一種解釋，特別是當這種解釋尚缺乏實徵驗證的支持，即聲稱完全可以解釋該事件時，就會對我們造成阻礙。

儘管，有時我們暫時不能找到對一個特定現象的系統解釋，但是不斷尋求概念化的加工系統仍然在不停地工作，企圖把無意義的理論強加於原本隨機的資料。心理學家曾對此現象進行實驗研究。研究者在一種實驗情境中，要求受試者觀察在許多不同維度上變化的一系列刺激物，並告訴他們其中的一些刺激物是屬於一類，其他的則屬於另一類，受試者的任務是猜測每個刺激物屬於哪一類。實際上，研究者是把刺激物隨機歸類，因此，除了隨機性，並沒有任何其他規律。但是，受試者很少敢做隨機猜測，相反地，他們通常處心積慮的虛構一個複雜的理論，以解釋他們對這些刺激物的分類。

同樣的，各種各樣的「預謀論」（conspiracy theories）通常也需要搞出一層又一層的複雜計謀，以

幫助那些積極尋求一個穩定解釋的預謀論者，把原本是隨機出現的一連串獨立事件，說成是有預謀的相關事件。例如，許多金融分析者的思維方式就體現了這一謬誤。他們通常對股票市場價格每一次小小的波動都作出精心的解釋，實際上，這種波動大多只是隨機波動而已（Malkiel, 1999; Taleb, 2001）。然而即使大量的證據表明他們做不到，股票分析師依然總是讓人感覺他們可以（或許他們也相信自己可以）「打敗市場」。如果你購買了包括在標準普爾指數（Standard and Poor's Index）中的五百大企業之股票，然後放著不管它（我們可稱之爲傻子策略），經過一九七〇年代的整整十年，獲得的報酬率會比華爾街80%的股票經紀人爲他們的顧客所賺的還高（Malkiel, 1999）；如果是在八〇年代的十年裡，你可以打敗華爾街三分之二的經紀人（Malkiel, 1999; Updegrave, 1995），也可以打敗80%那些每年要訂閱者花五百美元才看得到的財經通訊刊物（Kim, 1994）。

但話說回來，我們要如何看待那些確實打敗了傻子策略的經紀人呢？你可能會問，這是否表示他們具有某些特殊的才能。要回答這一問題，讓我們先設想下面這個實驗：有一百隻猴子，每隻猴子以手中的十支飛鏢，射向一面寫有標準普爾五百大企業名字的牆上，飛鏢投中的就代表猴子那年要買的股票。那麼，一年後牠們的業績如何？有多少隻猴子會打敗標準普爾五百大指數？你答對了！大概有一半的猴子。那麼，要你付錢給這一半打敗標準普爾五百大家的猴子，讓牠們下一年幫你炒股票，你願不願意？

有關財經預測的例子還可以幫助我們進一步說明，純粹隨機發生的一連串事件，是透過什麼邏

輯讓它看起來像是由可以預測的因素造成的。接下來的這個例子引自帕洛斯（J. A. Paulos, 1988）和佛雷德森（M. S. Fridson, 1993）。假想你收到一封信，告訴你有一份股票市場預測的電子報問世，這份電子報並不是要你訂閱，只是要求你試試照著他們的建議買股票，看看它的預測靈不靈。它告訴你ＩＢＭ的股票會在下個月攀升。你把這一電子報放在一邊，之後你注意到下一個月ＩＢＭ股票果眞漲了。如果你讀過一本與本書相似的書，可能就知道這並不代表什麼，你只會記下這是一個僥倖的猜測。後來你又收到另一份同一家投資諮詢公司的電子報，說ＩＢＭ會在下個月下跌，當股票確實下跌時，你又一次記下這是一個僥倖的猜測，但這一次你可能就有些好奇了。當這家公司寄來第三份電子報，預測ＩＢＭ下一月份會再次下跌時，你發現自己會更關注財經版的新聞，你發現這份電子報又一次作出準確預測，ＩＢＭ這個月確實又下跌了。當來自這家公司的第四份電子報說ＩＢＭ下月會攀升，而也確實攀升時，你就很難不會想，這份電子報還眞有點本事——很難擺脫想花29.95美元訂閱一年的衝動，唯恐這個電子報會在第四次免費贈閱之後就不再寄來了。

是的，如果你不知道其中的祕密，確實很難擺脫這種衝動。但是，如果你能想像，有個人在一個簡陋的地下室辦公室內，下週要發出一千六百份電子報，其中八百份預測ＩＢＭ下月攀升，八百份預測下跌，然後按電話簿上找到的一千六百個地址寄出去。那麼，一切就不足以爲奇了，當ＩＢＭ在下個月眞的攀升了，公司就繼續把電子報只發給上月接收到正確預測的八百位客戶（當然，其中還是四百份預測攀升，另外四百份預測下跌）。然後，你可以繼續想像，這個「氣鍋房」——可能與電

話行銷的炒作騙子共謀，利用電話進一步跟進騙局——正在向第二週接收到正確預測的四百位客戶發送第三個月的預測通訊。接下來，對了，你就是連續四次收到正確的隨機預測資訊的一百位幸運者中之一（多麼令人激動呀！），同時也是那一百個開始相信該通訊的準確度，從而會被進一步要求付29.95美元訂閱它的幸運兒之一。

知道內情之後，這是一個多麼可怕的騙局啊！確實是這樣。但是，這比「受人尊敬」的財經雜誌或電視節目展示的「股票經紀人連續四年打敗多半對手」的情節要好得多。請回想一下猴子射鏢的情境，設想這些猴子是財務經紀人，年復一年的幫人選股。很明顯地，第一年，牠們之中有50%會打敗對手。第二年，這50%的人中又有一半——按隨機機率計算——打敗牠們的對手，即只剩25%。之後第三年又有一半——按隨機機率計算——打敗對手，即只剩下總人數的12.5%。最終到第四年，又只有一半（總猴數的6.25%）能打敗對手。因此，一百隻猴子中只有六隻能達到財經節目和報紙所說的「在連續四年裡打敗其他的經紀人」的驕人成績。那麼，既然這六隻猴子打敗了其他猴子（正如我們前面看到的，也就是打敗了大多數現實生活的華爾街經紀人，參見Malkiel, 1999），牠們當然可以理直氣壯地在《華爾街一週》電視節目中亮相囉，你不覺得嗎？

◎解釋偶然：錯覺相關和控制的錯覺

人們喜歡對偶然事件做解釋，在心理學家研究的一個稱爲錯覺相關現象（illusory correlation）中

看得最清楚。當人們相信兩類事件通常應該一起發生時，就會認爲兩個事件一起出現的頻率是比較高的，即使它們的同時出現是隨機發生的，而且不比任何其他兩個事件同時發生的次數多。總之，人們甚至在隨機事件中，也傾向於見到他們預期的相關（Nisbett & Ross, 1980; Stanovich, 1999, 2004），他們看到了並不存在的結構。

有控制的實驗研究結果顯示，當人們預設有兩變數相關的想法時，即使在兩個變數根本毫無關係的資料中，他們也可以找到關聯（參見 Broniarczyk & Alba, 1994; King & Koehler, 2000; Stanovich & West, 1998）。不幸的是，這一現象在現實世界中也廣泛出現，以致給人們的生活帶來許多負面影響。例如，許多心理治療從業員一直相信羅夏克測驗（Rorschach test）是有效的。這個著名的墨跡測驗要求受試者對一張白紙上的墨跡做出反應。因爲墨跡沒有結構，所以測驗的發明者認爲，人們會以自己對模糊情境的定型反應對這些墨跡做出反應，從而揭露其「潛藏的」心理特質。這種測驗也稱爲**投射性**（projective）測驗，因爲發明者認爲受試者對墨跡的反應，應該是他們潛意識的心理、思維和感受的投射。

然而問題是，沒有任何證據表明羅夏克測驗作爲一個投射測驗，提供了任何額外的診斷價值（Dawes, 1994; Garb, Florio, & Grove, 1998; Lilienfeld, 1999; Lilienfeld et al., 2000; Shontz & Green, 1992; Wood, Nezworski, & Stejskal, 1996）。對羅夏克測驗的信心，可以說是源於錯覺相關這一現象。臨床心理醫師從病患的反應樣式中找到關聯，是因爲他們相信這些關聯的存在，而不是因爲眞的觀察到。

心理學家海曼（Ray Hyman, 1981）討論了人們喜歡在原本沒有關聯的事件中尋找關聯的傾向：

> 為了理解世上事物，我們必須運用自己原有的知識和預期。在大多數情境中，利用我們的知識背景和記憶，可以令我們正確地解釋對世界的觀察，並為這些解釋提供必要的推論。但這一有力的機制，在根本沒有資訊需要傳遞的地方就會走岔，我們會錯誤地、不斷地追求原本沒有意義的意義，而不是找出隨機性造成的噪音。（p. 96）

在我們的生活中，許多人與人之間的交往有很大的偶然成分在裡面：互不認識的男女約會最終促成了婚姻，取消的約談導致就業機會的喪失，誤了班車而遇到高中的老同學等等。認爲生活中每一件偶然的小事都需要認眞解釋的想法是錯誤的。但如果偶然事件確實產生重要的後果，人們不免要建構一些複雜的理論以解釋它們。

試圖解釋偶然的傾向，可能源於我們想相信自己可以控制這些事件的深切渴望。心理學家蘭格（Ellen Langer, 1975）研究了**控制的錯覺**（illusion of control）這一概念。它是指，相信個人的能力足以影響偶然事件的結果。在一個研究中，兩家不同公司的兩個員工賣樂透彩券給同事，對其中一些同事，他們只是簡單的給每人一張，對另一些同事則讓他們自行抽選一張。當然，在隨機抽獎的事件中，無論樂透彩券是有選擇權還是硬性派發的，中獎率都一樣。但是第二天，當這兩個員工試圖向這些同事買回那些樂透彩券時，有選擇權者對樂透彩券的要價是硬性派發者的四倍。在另外幾個實

驗研究中，蘭格證實造成這一結果的原因在於，人們不能接受個人的能力無法對偶然事件產生影響這一事實。

這一控制的錯覺在美國各州發行的樂透買賣中都有出現，顯示它是相當普遍的現象。這自然爲那些教導人們如何「戰勝」樂透的僞科學書籍的供應商開了方便之門，他們蜂擁而入去騙錢（Clotfelter & Cook, 1989）。這類書之所以暢銷，是因爲人們不懂得什麼叫隨機性。事實上，美國各州是到一九七〇年代中期才爆發購買樂透的熱潮，導火線就在於新澤西州發明了一種新的樂透彩券販賣方式：讓購買者自行刮彩券，或在彩券上挑自己要的數字（Clotfeler & Cook, 1989; Thaler, 1992, p. 138）。這類的參與性遊戲（participatory games）正是利用了當時蘭格剛研究出來的控制錯覺現象：人們錯誤地相信他們的參與行爲能夠決定隨機事件，讓他們有多點機會中獎。

一些心理學家則研究另一個與此相關的現象：**公正世界假設**（just-world hypothesis），它是指人們傾向於相信自己生活在一個公正的世界裡，每個人都得到他應得的（Lerner & Miller, 1978）。研究者證實，公正世界假設中有一個信念，會導致人們對因偶然造成的不幸事件之受害者不予同情，認爲他們是罪有應得。傾向於爲偶然事件尋求解釋，正是產生不同情這類受害者現象的原因，人們很難相信一個完美無邪或道德修養好的人，會因爲偶然事件而遭遇不幸。我們當然想相信好人有好報、惡人有惡報；但偶然卻是完全不帶偏見的，所有人都有同等的機率得到好的或壞的偶然事件。

一旦將公正世界假設這一信念推到極端，會產生一些有害或不人道的教條。從一九八〇年代早

期美國教育部的一個官員陳述的觀點背後之邏輯，就可見一斑。他說，殘障人士「錯誤地認爲生命的巧合使他們隨機地受到懲罰，其實不是這樣的。沒有什麼發生在一個人身上的事，不是他本人在生長過程的某一點上招致的……這句話聽起來好像很不公平，但是一個人所處的外部環境，確實與他內在心靈的發展一致。」（Gilovich, 1991, p. 143）正如基勒維奇指出的：「這眞不該是一個想進入教育部高層的官員應有的哲學，教育部本應是負責給予殘障人士同等的教育機會才對。」（pp. 143-144）——但這正是拒絕承認偶然這一因素在作用的不可避免的結果。

體現在公正世界假設中，對偶然的錯誤理解，也可以用來理解許多導致我們見到錯覺相關的不正確的民間信念。例如，我們在第六章提到，盲人有非常敏銳的聽覺，就是一個不正確的信念。然而，這個民間信念之所以流傳久遠，可能是因爲人們想看到這個關聯，讓他們覺得盲人得到應有的補償（參見 Niemeyer & Starlinger, 1981; Stankov & Spilsbury, 1978）。

偶然與心理學

在心理學中，研究者也有試圖解釋任何事件的習慣。他們不只對行爲中系統的、非隨機的事件提出理論進行解釋，任何一點細小的變化，也總會有人提出理論解釋。結果是造成不可證僞理論的泛濫，包括只是個人提出、未經公開化檢驗的理論和那些僞科學的理論。「心理歷史」研究法的奉

行者就經常犯下這一錯誤，一個著名人物生命中的每個細小的變化及轉折，通常透過精神分析學派的理論，都解釋成重大的心理歷史事件。這些心理歷史事件的問題，不在於它們解釋得太少，而是太多。此一研究方法的奉行者，很少承認決定一個人一生的因素有許多是偶然的。

對於想運用心理學知識的外行人來說，理解偶然這一因素的作用是非常重要的。受過正統訓練的心理學家承認他們的理論只能解釋人類行爲變化的一部分，而不是所有的，他們會坦然接受偶然因素的作用。但是，那個在《歐普拉脫口秀》出現的嘉賓（見第四章開頭），對每個個案及人類行爲的每個小節卻幾乎都能用他的理論提供一個解釋。我們對他的這一天分應該懷疑，而不是尊敬。眞正的科學家從不懼怕承認他們的無知。總之，評價心理學說法的另一實用法寶就是：在接受對一個事件的複雜解釋以前，先想一想，這一事件的哪個部分事實上是偶然因素的作用引致的。

◎巧合

對實質上是偶然的事件進行解釋的習慣，也導致我們對許多巧合（coincidence）事件的性質產生誤解。許多人認爲巧合需要特別的解釋，他們不理解巧合，即便只是偶然引起的，也是必定會發生的，巧合並不需要特別的解釋。

《韋氏新世界字典》（*Webster's New World Dictionary*）把**巧合**定義爲：「相互關聯或相同的事件意外地、不可思議地出現。」因爲在同一本字典裡把**意外的**（accidental）定義爲「因爲偶然而發生

的」，所以毫無問題的，巧合只是相關事物基於偶然的因素才同時發生的。不幸的是，許多人並不這樣解釋巧合，自動尋找事件之間的模式和意義的習慣，與巧合的「不可思議」這一性質結合在一起，使得許多人忘記可以用偶然這個因素解釋巧合；相反地，他們卻精心炮製一些理論以解釋巧合。下面我要講的這個故事你一定已經聽過不少遍了：「那天我坐在那裡，正在想好久沒有叫比爾叔叔到德州來玩玩了，緊接著電話鈴響，你猜怎麼著？是的，沒錯！正是我老比爾叔叔打來的，看來還眞有『心電感應』這玩意兒！」這是精心解釋巧合事件的一個典型例子。每天，大多數人都可能想到很多不同的、住在遠方的、認得的人，在我們想起他們時，有多少人可能會打電話來？幾乎沒有。因此，一年內我們可能想到數百個人，而其中沒有一個打電話來，但是我們並不介意，因爲這是常態。最終，在這數百個我們從未意識到的「否定事件」之後，一個罕見的事件發生了——當我們想到某個人時，他正好打電話來！我們可以發現，這是在數百件同樣的事件中難得出現的一個情況。然而難得的事件總是會出現的，而且出現純粹是偶然，其他解釋都是不必要的。

如果人們確實理解巧合的意思（一個因偶然而發生的不可思議的事件），就不會掉入尋找系統的、非偶然之解釋的陷阱。但事實正相反，巧合對很多人來說都不認爲是偶然的結果，而是需要解釋的。例如，許多人都聽過這一句話：「天哪！這眞是太巧了！**怎麼會**這麼巧？」爲此，馬克斯和卡曼（D. Marks and R. Kammann, 1980）建議大家，今後用一個比較中性的名詞**稀配**（oddmatch）表示某兩件事令我們感到非常奇怪地、不可思議地同時發生了。**稀配**這個詞並未具體指明兩件事同時發生是

基於偶然，還是基於系統的解釋。

認爲罕見的事不會發生，認爲稀配絕非偶然，這類錯誤觀念是我們之所以爲巧合事件尋求解釋的一個原因。這類錯誤信念之所以如此強烈，也是因爲機率有時是用兩件事的對比機率比例（odds）這一概念表述的，而其言外之意起著一定的作用。看看我們是怎麼說的：「啊！天哪，這事件是很不可能發生的！因爲它出現與不出現的對比機率比例是一比一百！」我們的表述方式強烈地讓人感覺這件事絕不會發生。當然，我們可以用另一種可能給人完全不同感受的方式表述同一件事：「在一百個同類的事件中，這樣的結果可能會發生一次。」這種表述方式強調，儘管這一事件是少見的，但是長時間來看，罕見的事終究一定會發生。簡言之，稀配會因純偶然而發生。

實際上，機率法則保證，隨著事件發生次數的增加，一些稀配會出現的可能性變得很大；也就是說，偶然事件的法則不僅讓稀配遲早會出現，而且長時間來看，幾乎可以說**保證**它一定會出現。看看馬卡斯和卡曼（1980）的例子，如果一次擲五枚硬幣都是正面朝上，你將認爲這是一個稀配、一件不可能發生的事。是的，它發生的機率是三十二分之一或0.03。但是如果你擲一百次，至少有一次全部正面朝上的可能性是多少？答案是0.96，也就是說，一百次中，這個罕見的稀配是**非常可能**發生的。

因爲美國很多州都設立了樂透，透過電腦或其他裝置隨機抽出中獎號碼，所以每週的搖獎活動累積下來，不可避免地會有極不可能的巧合事件出現，例如連續兩週中獎號碼相同，這一稀配的出

現雖然只會引起許多統計學家和行爲科學家的一兩聲哼笑，覺得是個奇妙的巧合，卻會引起大衆的抗議，認爲這表明搖獎過程有作弊或動手腳的嫌疑（參見 Clotfelter & Cook, 1993）。大衆之所以認爲有人作弊，起因於他們錯誤地認爲，這樣罕見或不大可能出現的事情，不可能只是偶然因素的結果。當然，統計學家哼笑的理由是，偶然恰恰是以與大衆的邏輯相反的方向起作用，如果樂透發行了一段長時間，連續出現相同中獎號碼的事必然會發生。

若干年前，蘭德（Ann Lander）宣傳了一系列有關亞伯拉罕・林肯（Abraham Lincoln）和約翰・甘迺迪（John Kennedy）總統之間令人「毛骨悚然」的巧合：

1. 林肯於一八六〇年當選總統；甘迺迪則於一九六〇年當選。
2. 林肯和甘迺迪都關注民權。
3. Lincoln 和 Kennedy 都是七個字母。
4. 林肯有一個叫甘迺迪的祕書，甘迺迪也有一個叫林肯的祕書。
5. 兩人都由叫詹森（Johnson）的南方人繼任。
6. 兩人都被有三個名字的人暗殺（John Wilkes Booth 和 Lee Harvey Oswald）。
7. Booth 和 Oswald 都持有不受人歡迎的政治理念。
8. Booth 在電影院裡射殺林肯然後藏在倉庫裡；Oswald 從倉庫中射殺甘迺迪然後藏身電影院。

當然，作爲巧合，這些事之間的關聯一點也不令人毛骨悚然。德州大學的電腦程式設計師李維（John Leavy, 1992）曾經搞了一個叫「令人恐怖的總統巧合競賽」，說明實際上在任何兩個總統之間找到像上述那樣的一份清單是多麼容易的事（參見 Dudley, 1998）。李維的文章中包括對威廉・哈里森（William Henry Harrison）和紥卡里・泰勒（Zachary Taylor）的比較，波爾克（Polk）和卡特（Carter），加菲爾德（Garfield）和麥金利（McKinley），林肯和傑克遜（Jackson），尼克森（Nixon）和傑佛遜（Jefferson），華盛頓（Washington）和艾森豪（Eisenhower），格蘭特（Grant）和尼克森，麥迪遜（Madison）和威爾遜（Wilson）等之間的相似之處。下面是加菲爾德和麥金利之間令人毛骨悚然的相似處：

1. 兩人都在俄亥俄州出生長大。
2. 兩人都是美國內戰的老兵。
3. 兩人都在衆議院任過職。
4. 兩人爲了保護美國工業，都支援金本位制和關稅保護制。
5. McKinley 和 Garfield 都是八個字母。
6. 兩人都被來自紐約市的副總統取代（羅斯福〔Theodore Roosevelt〕和亞瑟〔Chester Alan Arthur〕）。
7. Theodore Roosevelt 和 Chester Alan Arthur 都是十七個字母。
8. 兩個副總統都蓄鬍。

9. 兩人都在任期第一年的九月被暗殺。
10. 暗殺兩人的刺客Charles Guiteau和Leon Czolgosz聽起來都像外國人的姓名。

許多關於總統之間關聯的清單都很類似。總之，在一個人一生的數十年中，由於人際交往及各類事件的複雜性，在內藏著成千上萬事件的抽樣空間裡，如果任何兩個人之間找不到什麼相似之處，那才是讓人奇怪哪（Martin, 1998）。

懂得在什麼時候抑制對隨機事件編造複雜的解釋，是具有實際意義的。作家嘉萬迪（Atul Gawande, 1999）曾描述一九七三年贖罪日戰爭（Yom Kippur War）期間，認知心理學家卡尼曼的建議遭到以色列空軍部隊批評的故事。兩支飛機中隊出發了，回來後，一隊損失四架飛機，另一隊沒有損失。部隊希望卡尼曼調查一下，這樣的差異結果是不是有特別的因素。卡尼曼並沒有調查，他僅僅運用本章談到的邏輯告訴以色列空軍部隊不要浪費時間：「卡尼曼知道如果空軍長官真的調查，將不可避免地發現一些可測量到的差異，以致非要做一些舉措來彌補這一差異。」（Gawande, 1999, p. 37）。但是卡尼曼知道，任何找到的因素都極有可能是虛幻的——只是偶然波動的結果。

◎個人的巧合

在個人的生活中，稀配的發生往往因爲對我們具有特殊的意義，而更不易歸因於偶然。產生這

種傾向的原因很多，某些是動機性和情感性的，但還有一些是機率推理的失敗。我們通常不能意識到，稀配只是包括衆多「平凡」事件的大事件庫中一個非常小的部分而已。對某些人來說，稀配看起來好像經常發生，但是它眞的經常發生嗎？

想想看，如果我們現在分析一下你個人生活中的稀配，結果會告訴我們什麼呢（下面的例子引自 Marks & Kammann, 1980）？假定在某一天裡，你參與的不同事件有一百個，這個數目其實並不算多，想像一下我們是生活在多麼複雜的現代工業社會！實際上，一百個事件可能還少了點。你看電視、打電話、與人面談、商定去工作或去商場的路線、做煩人的家務、閱讀吸收資訊、在上班時完成複雜的任務，等等。所有這些事件都包含很多成分、而各成分又可以看成是一個個可分開記下來的小事件，加加起來，說一天中一個人涉及一百個事件，其實眞不算多。不過，我們就拿一百個事件作例子吧。

稀配是指其中兩個事件不可思議地連在一起。那麼典型的一天中這一百件事之間共有多少不同的、兩個事件的匹配呢？用一個簡單的公式算一下可以得到，一天中有4,950個不同的兩兩事件匹配，一年有三百六十五天，天天如此。我們知道，稀配是令人難忘的，比爾叔叔打來電話的那一天，你可能幾年不忘。假定你把在十年期間記得的所有稀配數出來，也許也就六、七件（或多或少，不同的人對稀配一詞有不同的標準）。這六、七件是從一個多大的平凡事件庫中出現的？每天 4,950 個兩兩匹配，乘以一年三百六十五天，再乘以十年，得到 18,067,500 個匹配。總之，十年中有六個

你認爲是稀配的關聯發生，但其他的18,067,494件事也可能發生稀配。所以，在你的生活中一個稀配發生的機率是0.00000033，這樣，在一千八百萬個事件聯結中，只有六個是稀配，其實應該不足爲奇。罕見的事件確實發生了，它們也的確少見，但是偶然這一因素保證它們一定會發生（請回想前面擲五枚硬幣的例子）。在我們的例子中，在你身上發生了六件奇事，它們可能是巧合：兩個相關事件基於偶然不可思議的同時發生。

心理學家、統計學家以及其他科學家都指出，人們認爲許多稀配的少見程度比實際上出現的頻率少得多。在一篇副標題爲「爲什麼巧合的長臂沒有看起來那麼長？」（1972）的文章中，迦納討論了許多稀配並不是像通常想像的那樣罕見這一事實。著名的「生日問題」是最好的例子。在一個二十三人的班級裡，有兩個人同一天生日的機率是多少？大多數人會認爲非常低。而實際上，二十三人的班級中，兩人同一天過生日的可能性是50％！而三十五人的班級，可能性就更大了（機率大於80％；參見Martin, 1998）。所以，因爲美國歷史上有四十二位總統，其中有兩位（波爾克和哈定〔Warren Harding〕）是在同一天出生（十一月二日），就沒有什麼了不起了。同樣的，有三十六位總統已過世，其中兩位（費爾摩〔Millard Fillmore〕和塔夫托〔William Howard Taft〕）死於同一天（三月八日），也不應令人驚訝，甚至還有另外三位總統（亞當斯〔John Adams〕、傑佛遜和門羅〔James Monroe〕）都死於同一天——而這一天竟然是七月四日，美國國慶日！你說奇不奇怪？其實不奇怪——機率使然而已。

這裡要表達的是，很多稀配其實並不奇特。心理學家范克和麥克格瑞格（Ruma Falk and Don Mac-

Gregor, 1983; Falk, 1989）研究人們如何判斷一件事是否值得驚訝。他們的一個重要發現是，在判斷巧合事件時，我們是很自我中心的：人們對發生在自己身上的巧合，比發生在別人身上的同樣巧合更驚訝。范克和麥克格瑞格（1983）推測這一結果的原因可能是「我們常常認爲自己是獨特的、與別人不一樣的，因而看不到發生在自己身上的事情，同樣也可能發生在別人身上。換句話說，我們可能不容易看到，在一個事件或情境中的自己，是可以換成另外一個人的。」（p.501）

接受錯誤以減少錯誤：臨床式與精算式預測

人們在試圖解釋世界上發生的事時，往往不願意承認偶然因素的作用，這一現象實際上降低了我們對現實世界的預測能力。在某個領域中，承認偶然因素的作用意味著研究者必須接受一個事實：我們的預測不可能百分之百準確，我們總是會犯錯。但有趣的是，承認達不到百分之百的準確度，實際上反而提高了整體預測的準確性。這聽起來好像有點矛盾，但事實確是如此：爲了減少錯誤，就必須接受錯誤（Dawes, 1991; Einhom, 1986）。

接受錯誤以減少錯誤這一概念，可以透過在認知心理學實驗室裡研究了數十年的一個非常簡單的實驗任務來說明（Fantino & Esfandiari, 2002; Gal & Baron, 1996）。這個實驗任務如下，受試者被安排坐在兩盞燈（一紅一藍）前，實驗者要求他們預測每次實驗測試時哪一盞燈會亮，受試者要參與很多

次這樣的測試，並按準確率給予一定的報酬。實際上，所有的測試都是在70%的次數亮紅燈、30%藍燈的條件下進行的，兩種燈以隨機順序出現。實驗過程中，受試者很快就感到紅燈亮的次數比較多，因此也就在更多的測試中預測亮紅燈。事實上，他們確實有大約70%的測試中預測亮紅燈。然而，正如前面討論的，受試者在實驗過程中逐漸發現並相信燈亮是有一定模式的，卻從沒想過每一次哪個燈會亮本身是隨機的。因此，爲了使預測百發百中，他們就在紅燈與藍燈之間換選，但保持70%的次數預測紅燈、30%預測藍燈會亮。受試者極少意識到，如果放棄力求百發百中，他們的預測會更好一些！爲什麼會這樣？

讓我們思考一下這個情境背後的邏輯。在以70：30的比例隨機點亮紅燈或藍燈的條件下，如果受試者70%的測試中預測紅燈會亮，30%預測藍燈會亮，他的準確率是多少？我們用實驗中間部分的一百個測試計算——因爲那時受試者已經注意到紅燈亮的次數比藍燈多，從而開始在70%的測試中預測紅燈會亮了。因爲在一百次測試中有70次亮紅燈，受試者在這70次中有70%的準確率（因爲受試者在70%的測試中預測紅燈會亮），也就是說，受試者在70次中有49次正確的預測；一百次測試中有30次亮藍燈，受試者在這30次中有30%的準確率（因爲受試者在30%的測試中預測藍燈會亮），也就是說，受試者在30次中有9次正確的預測。因而，在一百次測試中，受試者的正確預測是58次（49加9）。但請注意，這是多麼可憐的成績啊！如果在受試者注意到哪一盞燈亮得比較多後，就總是預測是那盞燈會亮——在本實驗中，就是注意到紅燈亮的次數比較多，就總是預測紅燈會亮（姑且稱

之為百分之百紅燈策略），一百次測試中將有70次正確的預測。雖然在藍燈亮的30次測試裡，受試者都沒有預測對，但是總準確率仍然高達70%——比在紅燈與藍燈之間換選，以追求百發百中的58%準確率高出12%。

然而，百分之百紅燈策略取得的較高準確率，是要付出代價的：必須放棄百發百中的雄心。（顯然的，雖然明知有時藍燈會亮，受試者還是預測紅燈，就表明他已經失去在藍燈亮的測試中命中的機會）。這就是我們所說的接受錯誤以減少錯誤！放棄在預測每一單一事件時不犯錯誤的想法，會讓受試者獲得更高的總體準確度。同理，如果我們要在預測人類的行為時有一定的準確度，有時就必須接受錯誤以減少錯誤，也就是，在依靠一般性原則作出比較準確預測的同時，也要承認我們不可能每個單一事件都預測準確。

但是，接受錯誤以減少錯誤，說來容易做來難。心理學這一領域裡，人們花了四十年時間研究臨床式預測（clinical prediction）和精算式預測（actuarial prediction）這一課題，就是有關「做起來難」的最好例子。**精算式預測**是指依據統計資料分析出來的群體趨勢所作的預測。在本章一開始所討論的群體（也就是總體）預測就是屬於這種。一種簡單的精算式預測是，針對凡是具有某種特徵的所有個體，作出相同的結果預測。所以，預測不吸煙者的壽命是七七．五歲而吸煙者是六四．三歲，就是一個精算式預測。如果我們考慮的群體特徵不只一個（運用第五章談到的複雜相關技術——尤其是**多元迴歸**）將令我們的預測更準確。例如，預測吸煙、肥胖且不作運動者的壽命是五八．二歲，就

是在一個多變數（吸煙行為、體重和運動數量）基礎上的精算式預測，這樣的預測總是比單變數更準確。精算式預測在經濟學、人力資源、犯罪學、商業與行銷學，以及醫學科學等領域都很常見。

在心理學的許多分支領域，如認知心理學、發展心理學、組織心理學、人格心理學與社會心理學中，其知識都是透過精算式預測來表達的。相反的，有一小群執業的臨床心理學家，則聲稱他們可以超越群體預測，對個別個體作出準確的預測，這種預測稱爲**臨床式預測**或**個案式預測**（case prediction）。與精算式預測對比，臨床式預測是這樣的：

> 職業心理學家聲稱，他們能對個體進行預測，從而超越對「一般人」或不同類別的人所進行的預測……職業心理學家與心理學研究者的不同在於，他們能把人作為獨特的個體來理解，而不只是群體的一部分，之所以只能作群體理解，是因為統計的概化必須是針對一群人做的。他們自認為能夠分析出，在一個**個體**的生活中，到底是「什麼導致了什麼」，而不只是說什麼樣的人「總體而言」是這個樣子的。（Dawes, 1994, pp. 79-80）

臨床式預測應該可以是精算式預測的一個有用的補充，但問題是，臨床式預測完全不準確。

要證明臨床式預測有效，一個臨床醫師與病患接觸的經驗，以及他對病患提供資訊的運用，應該會給我們一個比較好的預測。這個預測一定要好過我們只是運用對這些病患先進行簡單的量化編碼，再把這類資訊放在一定的統計程式中，以找出如何組合這些資訊，才能讓我們更精確地預測研

究的結果。總之，臨床式預測是有效的這一觀點認爲，臨床心理學工作者的經驗，會使他們超越精算式預測研究所能提供的總體相關資訊。既然講得這麼明瞭，那麼，我們就很容易檢驗這個觀點的正確性。很不幸，經過檢驗，此一觀點證明是錯的。

對臨床式與精算式預測的比較研究，結果具有驚人的一致性。自從米爾的經典著作《臨床式預測對統計式預測》(*Clinical Versus Statistical Prediction*) 於一九五四年出版以來，四十年間有一百多個研究表明，在每個進行過研究的臨床式預測領域（心理療法的效果、假釋行爲、大學生畢業比例、電擊治療的療效、累犯問題、精神病住院治療期的長短等），精算式預測幾乎都優於臨床式預測（如 Dawes, 1994; Dawes, Faust & Meehl, 1989; Faust, Hart, Guilmette, & Arkes, 1988; Goldberg, 1959, 1968, 1991; Leli & Filskov, 1981; Ruscio, 2002)。

在多個臨床領域中，研究者給予一個臨床心理醫師某位病患的資料，讓他預測這個病患的某些行爲。同時，研究者也把同樣的資料量化，並用一個統計程式分析及預測同樣的行爲，之後再看看到底哪一種預測比較準確。後者是以由先前相同研究找出的精算關係爲基礎來進行的。這些比較研究的結果毫無例外地顯示，統計程式是勝利者。這就表明，精算式預測比臨床式預測準確。事實上，即使是在臨床心理醫師可以得到的資訊比精算方法所用的資訊更多的情況下，精算式預測仍然更準確。也就是說，臨床心理醫師除了擁有與精算式預測一樣的量化資訊以外，還持有與病患個別接觸和訪談等個別化的資訊，但這些都無法讓他的預測像精算式預測那麼準確：「即使擁有資訊優

勢，臨床判斷仍然不能超越精算方法；實際上，擁有更多資訊，一點也不能彌補兩個方法之間的差距。」（Dawes et al., 1989, p. 1670）原因當然是方程式將各種資訊資料，以增加預測力爲目標作既準確又穩定的優化整合。在優化和穩定這兩個因素上，精算方法就遠遠優於臨床心理醫師透過非正式方法所蒐集的資訊了。

有關臨床式—精算式預測的研究文獻中，還包含另一種比較的方法：給臨床心理醫師由精算方程式得來的預測結果，讓他根據自己與病患接觸的經驗，對這一預測加以調整。結果，在臨床醫師對精算式預測作出調整後，實際上非但沒有增加，反而降低了預測的準確度（參見 Dawes, 1994）。這裡我們又看到一個不能接受錯誤以減少錯誤的絕佳例子，與前面所述的紅藍燈預測實驗非常類似。受試者不顧燈亮次數多少這一精算資訊，因而不採用每次都預測紅燈會亮的策略（可以獲得70%的正確率），而是在紅燈與藍燈之間換選，以力求每一次預測都正確，結果卻反而使正確率減少12%（只有58%的正確率）。同樣的，這裡所述的研究中，臨床心理醫師相信，他們的經驗應該可以給自己一些「臨床洞見」，從而作出比定量資料更好的預測。實際上，這些「洞見」根本不存在，他們的預測反而比依賴公開、精算的資訊所做的預測要差。最後請注意，精算式預測的優點並不局限於心理學，它也擴展到許多其他臨床科學中——例如醫學中對心電圖的解讀（Gawande, 1998）。

有鑒於研究顯示精算式預測優於臨床式預測，米爾（1986）曾說：「社會科學中，沒有任何爭論能像這個一樣，在這麼大量、性質那麼不同的研究領域中得到如此一致、單向的結果。」（pp.

373-374）但令人發窘的是，心理學領域並沒有根據這一研究成果身體力行。例如，儘管大量證據表明，個人面談這一考試方法實際上是無效的，但在研究生入學與心理健康培訓班入班的挑選等方面，還是有許多人使用。臨床工作者也繼續用一些似是而非的理由，爲他們依靠「臨床直覺」合理化，而不依賴更有效的總體性預測。例如，道斯等人（1989）注意到：

> 廣為流傳的反精算論點或陷阱在於，認為群體統計不適用於預測單一個人或事。這種觀點濫用了機率的基本原則……。要保持邏輯上的一致性，反精算論者必須承認以下想法是對的：如果一個人被迫只能玩一次俄羅斯輪盤，但是允許他選擇一把鏜內裝有一發或五發子彈的手槍。由於這是單一事件，所以那個人選擇哪一把手槍就變得無關緊要了。（p. 1672）

此一觀點的一個類比是，問你自己對如下一個科學發現的反應是什麼：一個完成過多次類似手術的醫師，下次類似的手術成功率較高（Christensen, 1999）。現在有個醫師A，他常做某一手術，失敗的可能性很小；另一個醫師B沒做過這種手術，失敗率可能很高，請問，你願意讓哪一個醫師爲你動手術？如果你相信「機率在單一事件中不適用」，就應該不介意讓醫師B動手術。

道斯（1994）討論了對精算式預測的另一反對意見，認爲它是「非個人化的」。他指出是否個人化其實無關緊要。我們在臨床心理學、醫學治療或財經諮詢等領域，爭論的是哪一種方法的預測最

好。道斯推測說，大多數人會選擇一種有效但非個人化的治療，而不理會一個無效但是個人化的治療：

> 對於感染ＨＩＶ（愛滋病）的人來說，有什麼會比非個人化地發明一種ＨＩＶ治療法更具個人意義及重要性？對於研究生入學申請者、假釋犯、心臟病病患……又有什麼比非個人化的統計預測技術的進步更具個人意義及重要性？關鍵在於任何一種方法是否能讓自己的在追求事業的過程中減少失敗，讓假釋犯在假釋期間減少再犯，讓醫學治療更有療效，等等。（p. 182）

承認在諸如預測心理治療療效等研究領域裡，精算式預測都優於臨床式預測，並無損於心理學的聲望，因爲在如醫學、商學、犯罪學、會計學，甚至家畜鑒別等差異非常大的應用領域中，也都發現精算式預測優於臨床式預測（參見 Dawes, 1994; Dawes et al., 1989; Dowie & Elstein, 1988）。儘管總體上說，心理學不會因爲這些研究結果而有什麼損失，但是對一些必須以「專家」身分從事臨床實務（例如在法庭上做證）、讓病患覺得他們有獨特臨床個案知識的工作者，不但聲譽會大損，甚至收入也會降低。然而，正如麥佛爾和特瑞特（R. M. McFall and T. A. Treat, 1999）在一篇論述臨床評估之價值的論文中提醒我們的：「我們臨床醫師試圖評價和預測的事，本質上是機率性的。這意味著我們不能期望大自然會如此聽話，讓我們有百分之百的把握來預測單一事件。相反地，我們最好的期望也

只能是確定一系列可能的結果，然後估計這些結果出現的相對可能性。從這種機率的角度看，傳統臨床評估內含的那種對獨特、不常出現的事件做出準確預測的理想化目標，其實是一個幻想，反映了我們的無知和／或自大。」（p. 217）

實際上，如果我們養成「接受錯誤以減少錯誤」的習慣，心理學和整個社會都將受益。在試圖對每個不尋常的單一事件做出獨特的解釋時（就我們目前的知識情況來說，獨特的解釋也許根本不可能），我們常常因而喪失對更常出現的平常事件的預測能力。請大家再次回想紅藍燈的實驗，在所有出現機率較小或較少出現的不尋常事件（亮藍燈）時，「百分之百紅燈策略」的預測都是錯的。如果我們把注意力放在出現機率較小的事件上，而採用「70%紅燈、30%藍燈策略」，結果會怎樣呢？我們會在30個不尋常事件中預測到9次（30×0.3），代價是損失了正確預測到21個常見事件的機會，從而把對紅燈有70次正確預測的結果，降爲只有49次（70×0.70）。臨床領域中的行爲預測也可以用同樣的道理，虛構每一事件的複雜度的解釋，可能確實能抓住一小部分不尋常事件——但代價是損失了對大多數事件的正確預測。簡單的精算式預測，正是對大多數、常出現的事件之預測做得比較好。嘉萬迪（1998）指出，醫學領域也同樣需要學習「接受錯誤以減少錯誤」這個道理。他認爲在醫學裡，強調直覺、個別化的治療方法「是有缺陷的——我們試圖承認及考慮人的複雜性，非但沒有減少錯誤，反而引來更多的錯誤。」（p. 80）

道斯（1991）講過一個有趣的故事，在他執教的大學，有人發起的一個罪犯教育專案，其中包

括與罪犯會談。那個罪犯原來是個酒保，額外還替人收賭債度日（必要時以槍威逼欠債者）。一天，他被派往另一州——奧瑞岡——收一筆錢。當他到達該州，把那個欠債者逼到街角，並掏出手槍要進行威逼時，警察從隱密處跳出來，當場逮捕他。向道斯講完他的故事之後，這個罪犯自信的說，他再也不會入獄了。道斯問他爲什麼那麼確信自己再也不會被捕？罪犯回答說：「因爲我再也不會去奧瑞岡州了！」

這個罪犯的思維有什麼錯呢？如果我們忽略事情發生的背景，接受他的邏輯，他的結論其實並不壞，正像道斯（1991）寫的：

> 當我講完這個故事時，人們都大笑起來，但從非機率的意義上來看，罪犯的推論並不是那麼壞。他很多次用手槍討債都沒有被捕，而他只到過奧瑞岡州一次。因而相信一定是奧瑞岡州有什麼特別的地方「導致」他被捕，是完全合乎歸納推理的原則。（p. 245）

此處的問題不在於推斷本身，而在於過分集中精力地想解釋一個特殊事件。罪犯想解釋爲什麼他會在奧瑞岡州被捕——爲什麼他會在那個特殊的時刻被捕。正由於過分關注他是在什麼特殊情境下被捕，使他得出那個非常荒謬的結論。這個罪犯應該運用機率性思考，每一次他用手槍討債時都有被捕的可能性（也許一百次中有三次）。罪犯討了很多次債都沒有被捕，這次他被捕了，那只是三次中本來就會有的事件。一個總體的趨勢（他每次用手槍討債就會有多大多大的可能性被捕）解釋了他爲

什麼會被捕。這件事發生在奧瑞岡州，可能只是剛好，根本沒有什麼特別的。

在這個例子中，透過類比，我們看到忽略較準確的精算資訊而依靠臨床資訊所帶來的損害。罪犯就像是一個忽視機率而依靠獨有的個案知識做預測的臨床工作者，只不過，他依靠的是對自己特殊經歷的總結知識作預測。罪犯對其現實世界經歷所累積的知識——亦可稱爲個人的洞察力和直覺——使他做出如果在奧瑞岡以外地方用手槍討債就不會被捕的臨床式預測。相反的，精算式預測會說，如果他繼續用手槍討債——在任何地方——他最終都會被捕（如果他每次討債不被捕與被捕的對比機率比例是九比一，那麼，他在八次討債中被捕的機率已經超過50%）。在奧瑞岡州發生這一事件也許很特殊，也許不特殊，但道斯（1991）反駁：

> 當然，在那被捕事件發生之前，確實有一系列的事件是在其他事件發生之前沒有出現的，但是否就一定要把這一系列事件視為被捕的「原因」呢？這樣的歸因有什麼用呢？實際上，試圖尋找這類的原因，非但沒有好處，還可能會助長自毀或危害社會行為的延續，因為一旦確定，「為什麼」這樣的行為在某種場合會造成不良後果而其他場合不會，我們就會覺得只要避開那個特殊的場合，重複該行為是很安全的。（p. 245）

這一討債人的例子，其實與一些眞正的研究結果非常相似。確實，有關精算式預測優於臨床式預測的研究結果，在臨床或個案知識最能發揮其價值的地方——有關我們自己的知識——也已得到

證實。瓦隆、格里芬、林和羅斯（R. Vallone, D. W. Griffin, S. Lin, and I. Ross, 1990）研究大學生預測未來一學年中會發生在自己身上之事件的能力，受試者被要求預測一個特定事件是否會在他們的生活中發生，並就該事件發生的可能性給出一個確信度的判斷。這些事件包括學業上的（例如「今年我將有一門課不及格」）、友誼方面的（例如「我將有個女朋友／男朋友」）、與家庭關係相關的（例如「我每個月會給父母打兩次以上的電話」），和活動方面的（例如「我將去舊金山觀光」、「我將加入一個學校社團」）。瓦隆等人（1990）發現，總體而言，學生對他們的預測過分自信：他們確信自己能預測未來行爲的程度，遠比應該確信的大。這個研究重驗了一個著名的知識預測實驗所發現的現象：人們傾向於過分自信（Bornstein & Zickafoose, 1999; Brenner, Koehler, Liberman, & Tversky, 1996; Griffin & Tversky, 1992; Schraw & Nietfeld, 1998; Yates, Lee, & Bush, 1997）。

當瓦隆等人（1990）檢視受試者預測自己未來行爲所犯的錯誤時，發現一個有趣的現象：這些受試者只有在違反某一特定事件發生的基本機率時，才會做出錯誤的預測。基本機率是指把所有受試者看作一個整體，用某一特定事件在他們身上發生的總人數計算該事件發生的機率（發生人數除以總樣本人數）。有些事件發生的基本機率比較高（樣本中的大多數人都會發生），另一些事件較低（樣本中很少人發生）。幾乎所有學生犯的錯誤——以及所有他們顯示的、錯誤的過分信心——都是在違反這一基本機率的情況下出現的。違反基本機率常會表現在受試者說出這樣的話：「啊，這雖然可能會發生在大多數人身上，但**不會**發生在我身上。」或者「這雖然在大多數人身上可能都不會發生，

但它會發生在我身上。」實際上，如果受試者以基本機率預測自己的行爲，而不是把自己視爲例外，他們會預測得更準確。由於認爲自己獨具有關自己的知識，因此深信對自己行爲的預測會更準一些，實際上，這種自信反而使他們的預測更不準確。

這個研究又是一個臨床式—精算式預測的例子，也是一個不遵循「接受錯誤以減少錯誤」的例子。正如瓦隆等人（1990）指出的：「有關行動者的知識不管有多少——即便是行動者自己及其過去經歷的知識——也無法讓預測者像明智地運用精算的基本機率知識那樣，成功地提高預測的正確性和改進精密度。」（p. 590）

華格納和凱倫（W. A. Wagenaar and G. Keren, 1986）闡述了，對個人自我知識的過分自信以及對統計資訊的大打折扣，如何使宣傳繫安全帶駕車的交通安全推廣活動失敗。因爲人們總是認爲：「我和別人不一樣，我開車很安全的。」問題是，85%的人都認爲自己的技術「比一般駕駛高明」（Svenson, 1981）——這顯然是很荒謬的。

相信「統計數字不適用於單一個案」所造成的類似謬誤，是賭徒惡習難改的重要因素。華格納（1988）在他的賭博行爲研究中總結道：

> 從我們和賭徒的討論中可以非常清楚的看到，賭徒大體上都能意識到賭博的長期後果不利於自己。他們也知道最終輸的會比贏的多，而且永遠都是這樣。但他們卻不能把這些

總體的統計考慮應用到下一局、下一小時或下一個晚上。滿筐滿簍的祕笈、祕訣、靈感等讓他們覺得，統計學在下一局或下一小時裡是派不上用場的。而他們相信自己還是可以預測下一局的結果。（p. 117）

華格納發現，強迫性賭徒有很強烈的傾向不去「接受錯誤以減少錯誤」。例如，二十一點牌局的玩家，普遍存在拒絕使用一種叫**基本率策略**——一種保證可以把莊家的贏率從6%或8%減少到不足1%的策略——的現象（參見 Wagenaar, 1998，第二章）。基本率策略是一個長期性的統計策略，必須在一段長時間使用的情況下才會見效。強迫性賭徒之所以拒絕不用，是因爲他們相信「有效的策略應該每一次都有效」(p. 110)。在華格納研究中的賭徒「總是說像基本率這樣的統計系統所教授的一般性策略是不會有用的，因爲它們忽略了每一具體情境的獨特性」(p. 110)。拋棄能保證他們少輸上千美元的精算式策略後，這些賭徒徒勞地轉而追求那些建立在每一具體情境之獨特性基礎上的臨床式預測。

當然，在這裡有關臨床式－精算式預測研究文獻的討論，並不表示個案研究在心理學中沒有作用，請大家記住，這一章談的只是預測行爲這一特殊情境。回想一下第四章對個案研究之作用的討論，我們當時已經指出，在引發對重要的、需要進一步探研之變數的關注方面，個案資訊是非常有用的。而本章說的則是，一旦重要的相關變數確定，而我們要開始運用它們預測行爲時，測量這些

變數並使用統計方程式進行預測，是最好的研究方法及程序。首先，透過精算式方法可得到較準確的預測；其次，用精算式方法得出的預測是一種對所有人開放的公開知識——任何人都可以使用、修改、批評或爭論。相反地，使用臨床式預測，就等於要依靠個別權威的個人學歷及經驗，而他的評估——正因爲其聲稱是單一及個人獨有的——因此也不可能付諸公開的批評與監督。正如道斯所言：

> 一個專業心理學者會在法庭上聲稱：一個極具殺傷力的「臨床判斷」是基於他多年累積的經驗，是無法公開證實的，只能在諸如專業資歷、給予相似的證詞有幾年、態度品行等不相干的地方接受挑戰。相反的，一個統計模型是公開的，所以可以隨時在理性的基礎上接受挑戰。（p. 104）

摘要

偶然在心理學中的作用，常遭外行的大衆和臨床心理從業者的誤解。人們很難認清，行爲事件結果的變化有一部分是偶然因素造成的；也就是說，行爲的變化有一部分是隨機因素作用的結果。因此，心理學家不應該以自己能預測每個個案所涉及的個別行爲自許。心理學的預測應該是機率

的，是對總體大趨勢的機率性預測。

讓人覺得自己可以在個體層次上進行心理預測，是臨床心理學家常犯的錯誤。他們有時候會誤導他人認為，臨床學歷及經驗賦予他們一種對個體案例做準確預測的「直覺」能力。恰恰相反，幾十年來，值得一看的研究都一致表明：在解釋人類行為的原因方面，精算式預測（基於群體統計趨勢的預測）遠遠優於臨床式預測。目前還沒有證據表明，臨床直覺能預測一個統計趨勢是否會在某個具體特定的個案中出現。因此，當一個人要預測行為時，千萬不要把統計資訊擱在一邊。統計預測也正確地告訴我們，當一個人在預測人類行為時，錯誤和不確定性是永遠無可避免的。

第十二章　讓人瞧不起的心理學

雖然大衆對心理方面的話題，總是帶有濃厚的好奇心及神祕感，但是他們對這一領域及其所取得之成就的評價，卻多半是負面的。心理學家都意識到這個形象問題，但大多又感到束手無策，於是他們只能不予理會，假裝這個問題不存在。然而，這樣做是一個錯誤。當大衆傳媒在形塑大衆對事物的感知上變得愈來愈具影響力時（例如，對那些不讀書看報的一般老百姓而言，小說式的電視紀實劇情片裡說的，就已經是眞正發生的歷史了），不理會心理學的形象問題，只會使問題愈變愈糟。

羅尼・丹佐菲（Rodney Dangerfield）是美國一個廣受歡迎的喜劇演員，他有一句抱怨現在已經成爲他的正字標記：「我老是被人瞧不起！」從某種意義上說，這也正是心理學在一般老百姓心中的地位。

本章就是想談談爲什麼心理學會像羅尼・丹佐菲那樣，無法得到應得的尊敬。

心理學的形象問題

有關心理學形象問題的成因，有些我們在前面討論過了。舉例來說，第一章討論的佛洛伊德問題就是其一。佛洛伊德理論的不能證僞性，無疑促成人們對心理學的低評價。如果大衆認識任何一個心理學家，這個人不是佛洛伊德就是史金納。對他們理論的歪曲報導及傳播，一定會導致大衆認爲心理學是一門非常膚淺的學問。當一門學科中最有影響力的學者，被認爲是主張人沒有思想、人和老鼠沒有差別，這門學科在外界大衆心目中還有什麼尊嚴可言？當然，史金納並沒有否認人類能夠思考，他從動物身上發現的許多有關工具性條件反射的定理，已證實**的確**可以推廣到人類行爲上。然而，大衆對這些科學事實卻知之甚少。同樣地，對佛洛伊德理論的歪曲，也降低了心理學在大衆心目中的形象。

◎心理學和超心理學

除了佛洛伊德和史金納的研究之外，一般人對其他卓越心理學的研究幾乎一無所知。爲證實這一點，你可以到附近的書店看看可以買到哪一類心理學的讀物。你會發現，那些擺在書店賣的心理學讀物通常可以分爲三類。第一類是一些心理學的早期經典著作（佛洛伊德、史金納、榮格〔Jung〕、弗洛姆〔Fromm〕、艾瑞克森等），而這些著作幾乎已和當代心理學毫無瓜葛了。

在多數書店中可以找到的第二類讀物，是那些假冒心理學的僞科學書籍，裡面充斥著永無止盡的有關超常現象的記載及報告，如心靈感應、千里眼、意念移物、預知、轉世、生物規律、星運投射、金字塔力量、植物溝通、通靈手術等（Lilienfeld, Lohr, & Moirer, 2001）。書店中這類所謂心理讀物的大量存在，導致以及反映了人們對科學心理學的誤解。他們認爲心理學家就是證實這些超常現象存在的人。這種誤解對心理學而言無疑是一大諷刺。事實上，心理學與那些超常現象毫無關係。現代心理學有興趣研究的東西完全不包括那些超常現象，其中緣由很可能會讓許多人大吃一驚。

要說對超感官知覺和其他超能力的研究不能被認爲是心理學的一部分，可能會引起許多讀者的不滿。多個調查結果一致顯示，超過50%的大衆相信有超感官知覺的存在，並且對這一信念堅信不疑（National Science Board, 2001）。歷史研究和調查可以讓我們對大衆之所以如此熱中於這些信念看出一些端倪（Alcock, 1987; Bainbridge & Stark, 1980; Grimmer, 1992; Stanovich, 1989, 2004）。物質至上的文化削弱了許多人的傳統宗教信仰，這些人就開始尋找其他形式的神祕主義以填補這一空缺。像許多宗教一樣，許多所謂的超常現象也標榜轉世之類的說法。對部分人來說，這一來世的說法能滿足他們想超越現有生命極限的需求。這就可以說明，爲什麼當心理學家說用科學心理學研究方法無法證實超感官知覺的存在時，人們無法接受。聲稱心理學不把超感官知覺視爲一個可行的研究領域，總是讓那些信徒感到生氣，他們指責心理學家過於武斷的把一些研究課題摒除於心理學研究之外。如果心理學家不理會這些人提出的抗議，他們事實上完全沒有減輕大衆對心理學的誤解。心理學家的確應

該針對爲什麼這些抗議是沒有根據的，給予一個謹愼而淸晰的解釋。下面就是我們對這一誤解做出的解釋。

科學家並不是根據什麼經文來確定研究課題的，也沒有什麼宣言指出什麼是可以、什麼是不可以研究的。研究領域的出現、擴展或終止，依據的是理論與方法的自然淘汰規律。也就是說，那些能夠產出碩果的理論和實證發現的領域，就會吸引大量的科學家繼續研究；那些陷入死胡同的理論，或沒有可能重複驗證的實證領域就會被離棄。這種對理論與方法的自然淘汰過程，引導科學不斷地接近眞理。

例如，當代心理學不認爲超感官知覺是個可行的研究課題，就是因爲對它的研究一直無法累積正向的成果，因此，心理學家就不再感興趣。在這裡我要強調**當代**一詞，是因爲心理學家在先前確實有一段時間對超感官知覺領域懷著極大的興趣，只是在累積大量的負面證據之後才失去興趣。正如歷史展示的那樣，研究課題通常不是由政府某個權威機構宣佈停止研究的，它們只是在生存競爭的環境中被自然淘汰出局了。

在心理學領域裡，超感官知覺從來沒有被認爲是一個不能研究的課題。有關這一說法的證據，不但淸楚並且很容易取得（Alcock, 1990; Druckman & Swets, 1988; Hooft, 2000; Hyman, 1996; Marks, 2001; Milton & Wiseman, 1999）。過去有許多研究超感官知覺的論文發表在正式的心理學研究刊物上。那些在媒體上頻頻曝光的超心理學家，總喜歡讓人們覺得這是個嶄新的領域，借此以爲驚人駭世的新發

現問世了。實際上，事情根本就沒有那麼戲劇性。

對超感官知覺的研究，可以說和當代心理學的歷史一樣久遠，它並不是一個新的研究領域。在心理學文獻中，它也曾經像許多現在被認爲是可行的課題一樣認眞研究過。然而，在正式心理學雜誌上發表的、有關這一領域的許多研究結果，大多顯示它是子虛烏有的。在研究超過九十年之後，我們仍然無法在嚴格控制的情境下，成功地重驗任何超感官知覺。儘管過去幾十年來做了大量有關超感官知覺的研究，卻從來沒有一個研究可以達到這一簡單的基本科學標準。甚至許多超心理學家和其信徒也不得不承認這一點（參見 Alcock, 1981, 1990; Druckman & Swets, 1988; Krippner, 1977）。正是因爲這個原因，才令當代心理學家對此一課題失去興趣。

有時人們會受到誤導，認爲科學家否認超感官知覺的存在，是因爲它違反了當代普遍接受的、有關大自然的理論。從第一章和第二章討論的科學程序來看，很明顯地，這種說法是錯誤的。所有科學家都忙於推翻現在已被接受的、有關自然事物的理論，因爲，不斷透過改變和精化現有的觀點——但當然要在遵循關聯原則的大前提下——我們才能逐漸接近眞理。當一種新的現象與現有的知識結構衝突時，科學家會對此提出疑問並尋求其他可能的解釋。所以，心理學家不相信有超感官知覺，並不是因爲它與原有的知識累積矛盾，而是因爲根本沒有任何科學證據支持它的存在。簡而言之，沒有被證實的現象，自然不需要作任何科學的解釋（參見 Alcock, 1981, 1984, 1990; Hines, 1988; Humphrey, 1996; Hyman, 1992, 1996; Milton & Wiseman, 1999）。

現在，諷刺的事來了，心理學家在許多測驗超能力是否存在的大會上卻扮演非常重要的角色，重要性可能僅次於專業魔術師，這些魔術師當然是下最多功夫拆穿多數超能力表演其實都是騙局的功臣（Randi, 1986, 1987）。很多相關論述，以及質疑有超能力的重要書籍，也出自心理學家之手。

這個現象的諷刺性顯而易見。心理學作爲一門可能是目前對超感官知覺的存在問題最能給予精確評估的學科，在大衆的心目中，竟然正是與這些僞科學混爲一談、難分難解的學問！正如下面我們還會再詳細討論的，心理學常常會陷入這樣一個「裡外不是人」的雙重災難中，這裡所說的僅僅是其中一例罷了。那種認爲在心理學裡沒有什麼行規的想法，以及認爲這個領域缺乏科學標準以判定知識之可信度的信念，常常導致人們把心理學與超感官知覺這樣的僞科學聯結在一起。然而，如果心理學家有朝一日眞的能夠成功地讓大衆認識到這些僞科學的眞面目，那麼，由於在人們心目中心理學與僞科學總是聯在一起，這些證據又可能被視爲是說明心理學不是一門科學的鐵證！

◎自助式讀物

在書店裡常見的第三類心理學讀物，就是所謂的自助式讀物。當然，這類讀物也有許多不同種類（參見 Fried, 1994, 1998; Fried & Schultis, 1995; Paul, 2001; Santrock, Minnett, & Campbell, 1994）。有些書是勵志型的，目的是爲了提高人們的自我價值感和自信心；另一些書則試圖將某些熟悉的老生常談用新的表達方式包裝起來；只有少數（實在是太少了）幾本是由負責任的心理學家爲大衆撰寫的。還有許

多書，不包括在上述第二類中，爲了確保獨特性而聲稱自己發明了一些新的「療法」，不但可以治癒某些特殊行爲問題，還能滿足大眾的一般性需求（賺多一些錢、減肥和提高性生活品質是其中的「三大」），從而得以暢銷。這些所謂新療法的成效，很少建立在控制實驗研究的基礎上。如果作者是個臨床醫師，他們通常只是依靠個人經驗或少數的幾個病例，來支持自己的新發明。

許多行爲和認知療法，都曾透過相當嚴格的心理學檢驗程序以支持其有效性，但是有關這些療法的書籍卻很少出現在書店的書架上。電子傳媒的情況更糟糕，電台和電視台幾乎沒有任何正經八百的心理學報導；相反地，他們總是請來許多「僞療法」的散播者和愛出風頭的媒體名人上節目，而這些人與心理學這個領域根本風馬牛不相及。媒體之所以這樣做的主要原因是，正規的心理治療法從來都不會聲稱自己能令病患馬上痊癒，甚至不會擔保治療一定成功或誇大治療範圍（例如，「你不僅會把煙戒掉，而且生活的各方面都會得到改善！」）。同樣地，現在網際網路上也出現類似的情形——缺乏嚴格的同儕審核，在網路上推銷的治療方法通常都是不可靠的。

大眾大多不知道，出版業並沒有爲讀者就自助式讀物把關，提供品質保證。佛雷德森（1993）曾講述了一位大眾心理學讀物作家的故事。這位作家寫了許多有關就業方面的暢銷書。他在《富比士》雜誌（*Forbes*）上有個專欄，並且經常在一些電視訪談節目上出現。不久，人們發現這位作家並沒有像他本人聲稱的擁有博士學位（事實上，他得到的是一個未被政府認可的函授學校的博士文憑）。進一步的調查揭示，這位作家聲稱自己曾爲研究一個問題所做的八千次訪談中，有一些是編造出來

的。當人們質疑出售這些騙子所寫的、具誤導性的書籍背後之商業道德問題時，這位作家的一個出版商說：「我們出版商做的事，有99％是基於對作者的信任。」(p. 144) 閱讀本書的讀者，從今以後要對信任一詞附上一個重要的註腳：要具備有關現代科學驗證方法背後之邏輯的知識。

在美國書市中占相當比重的這些自助式讀物，使人們對心理學的一般看法產生許多不良影響。首先，像佛洛伊德的影響那樣，這些書使大衆搞不清心理學研究關注的焦點所在。舉個例子，雖然有相當數量的心理學家為肥胖、人際關係和性問題提供治療，並不斷地做研究，但這個數量比起自助式讀物中所說的少得多。這種誤解，也使得大衆以為大多數的心理學家都忙於關注異常行為的研究和治療。事實上，大部分的心理學研究是關注人類的正常行為。

除了引起對研究內容的誤解之外，自助式讀物還讓人們對心理學的研究方法和目的產生錯誤印象。正如第四章討論到的，心理科學並不認為幾個個案研究、見證敘述和個人經歷陳述，就能構成充分的實徵證據，用以支持某種療法的有效性。然而，這些方法卻正是大多數自助「療法」提供的支持證據之類型。自助式讀物因此誤導了社會大衆，使他們認為大多數的心理學理論就是在這一類證據的基礎上得出的。我們在第八章中已闡明，一個理論的驗證需要許多不同類型的證據支持，其中個案研究證據提供的支持是最薄弱的。視這一類證據為證實某一理論或療法的確鑿證據，是犯了根本性的錯誤。

◎食譜式知識

最後，自助式讀物使大衆混淆了心理學的目標和多數心理學研究所追尋的知識類型。心理學家夏佛（Leigh Shaffer, 1981）認爲這種讀物留給人們一個深刻的印象：心理學研究者追求的是**食譜式知識**。食譜式知識是指那些只知道如何使用某物，對背後運作的基本原理一無所知的、「知其然不知其所以然」的知識。例如，大多數人知道很多如何使用電話的知識，他們知道如何撥號、如何獲得資訊、如何打長途，等等。但許多人對電話運作背後的物理原理一無所知。他們不知道電話是怎樣實現通話的功能，只知道自己能夠使用它。這就是電話的食譜式知識。在我們的社會裡，許多有關技術性產品的知識都是食譜式知識。

當然，這也不完全是壞事。事實上，多數技術產品就是按照用戶對背後的運作原理一無所知的情況設計的。食譜式知識這一概念提供一條可以區別基礎研究和應用研究的途徑。基礎研究工作者尋找自然界的基本原理，而不考慮這些原理能否轉化爲食譜式知識；應用研究工作者則致力於將基本原理轉化爲一些實用的產品，這種產品只需要食譜式知識就可以使用了。

多數自助式讀物只提供關於人類行爲的食譜式知識，它通常是以下面的形式呈現：「你只要做X，就會變得更加Y了」或「做A，其他人會對你更B」。當然，如果這個食譜是正確的，做這些事也不一定就是錯的，許多正規的心理治療都提供大量的食譜式知識。然而，當人們錯誤地認爲，

所有心理學研究的終極目標就是提供食譜式知識時，問題就來了。雖然很多心理學研究者確實致力於將基本的行爲理論轉化爲可以應用的心理治療法、保健行動方案或有效的工業組織模式，但心理學主要還是一門發現行爲的普遍事實和理論的基礎學科。在這裡，我們看到爲什麼外界覺得心理學研究很「怪」的另一個原因：基礎理論的研究與應用研究有很大的差異。

如果有一個人走到分子生物學實驗室，詢問一位研究者頭痛時應當服用兩片還是三片阿斯匹靈，我們會覺得這個人很無聊。原因不在於分子生物學與止痛藥沒有任何關係，將來對止痛藥的研究可能會運用到這一領域的知識。這個人的問題很無聊，是因爲分子生物學家並不是開藥方的人，他不能解答你要吃兩片還是三片阿斯匹靈的問題。研究者關注的是有關生物物質在分子層次的基本資料，這些資料可能會爲許多不同領域提供食譜式知識，但發現基本資料和將這些資料轉化爲食譜式知識的人不大可能是同一個，轉化爲食譜式知識的方法，也會與最初發現事實的方法有所不同。

由於自助式讀物導致大衆相信多數心理學家是在追尋食譜式知識，心理學家實際所做的大量基礎研究，看起來就有些怪怪的了。赫奇讓受試者在一間黑暗的房間裡注視著一個小紅燈，到底與我們的現實世界有什麼關係？是的，表面上看確實沒有一點關係。赫奇是對人們的視覺系統如何適應黑暗的基本原理感興趣，這些基本原理最終會轉化成食譜式知識，告訴人們如何處理一些特定的問題，例如因維他命缺乏而導致的夜盲症。然而，這一轉化並不是由赫奇本人完成的，而且它要在幾年之後才到來。

因此，自助式讀物給大眾對心理學的感知帶來兩個不良的副作用。第一，這些讀物裡談論的問題，並不一定是當代心理學關注的焦點，反而是反映了消費者想閱讀的內容。心理學系的學生通常無法充分意識到，出版是一種商業行為，市場的力量決定了什麼樣的讀物應當放在書店的書架上。然而，科學的關注並不是這樣的。在所有的學科領域，尤其是心理學，科學家認為有生命力的念頭，與那些商業包裝、具有銷售力的念頭，完全是兩碼子事。

第二，自助式讀物讓心理學看起來完全是一種食譜式知識。這是完全錯誤的，它忽略了心理學領域所進行的大量基礎研究。

心理學與大眾傳媒

大多數心理學家本身已經意識到，這個學科的大眾形象問題已導致一些不良後果。在這裡，我們充分看到此一領域禍不單行的情況。許多心理學家非常關注，如何將他們的學科在大眾的腦海中與自助式讀物和偽科學劃清界線。這種關注使許多心理學家小心翼翼，避免對人類迫切需要解決的問題提出肯定的解決辦法，以免自己與那些自助式讀物作家或偽科學散播者同流合污。這種不願承認自己在某一社會關切的問題上有專門知識的做法，已經變成各大學的研究生院訓練心理學研究者的基本態度了。心理學研究是最具自我批評力度的學科之一，心理學雜誌充斥著各式各樣的論文，

針對各種研究方法的問題進行批評及警告，有些問題我們在前面幾章中都有提到（例如，從相關關係推斷到因果關係的陷阱，聚集證據的原則，以及檢驗其他可能解釋的重要性，等等）。

如此一來，心理學對外公佈研究結果時，所採取的保守做法，就根植於大多數心理學研究者的心中。心理學家也因此不願意對外聲稱，他們對於目前人類迫切需要解決的社會問題有任何決定性的答案。這樣做，當然往往是對的。人類行爲的問題錯綜複雜，而且研究起來並不容易。謹慎一點是好的；然而，當心理學與媒體打交道時，就不是那麼一回事了。這一保守做法產生一些不幸的後果。

大衆傳媒的獨特邏輯是，如果大衆對某特定的心理問題感興趣，媒體就會去製作一個故事——不管是否眞有故事可講。當一位科學家告訴一位記者：「對不起，那是個複雜的問題，還沒有相關的研究資料，所以我不想對此發表意見。」這樣的回答絕對阻止不了那位記者去別的地方繼續尋找答案，直到他找到爲止。他找到的，通常是一位不那麼保守的科學家（對心理學問題而言，他通常只要找到**任何一位**稱得上是「權威」的人即可），這位人士就一定不會羞於下結論了（Murray, Scwartz, & Lichter, 2001; Ruscio, 2000）。

凱利、洛頓和卡爾佛（I. Kelly, J. Rotton, and R. Culver, 1985）討論到這一災難性的「媒體邏輯」如何使月亮效應（lunar effect）變得家喻戶曉。所謂月亮效應，就是指相信月亮的盈虧會影響人類行爲，特別是異常行爲。凱利等人分析了三十七個不同研究的結果，發現不存在人類行爲受月亮效應

影響的證據。反而，他們發現「媒體效應」的影響很大，媒體效應就是指媒體能夠指鹿爲馬，使人們相信一些根本不存在的現象。凱利等人不客氣地挑明我們剛才討論的媒體邏輯：

> 報紙、電視節目和電台的脫口秀，總是偏愛那些宣稱滿月能影響行為的人……，當記者打電話給我們，如果我們確認的說「當滿月時，街上到處都是失去理智的人」，他們自然高興得多。不幸的是，我們不能那樣做。當一個科學家不能給一段能夠讓他們轉化為煽情標題的言論時，記者總能找到一個可以提供所需言論的所謂「專家」。（p. 133）

新聞學教授麥克道格爾（Curtis MacDougall）在他的著作《迷信和報業》（*Superstition and the Press*, 1983）中，給出另一個媒體邏輯的例子。有一位記者經常寫通靈能力方面的故事，當問及他本人是否相信眞有這回事時，這位記者回答：「我沒有必要相信它。我要的是兩個有博士學位的人告訴我事情是如何如何發生的，我就可以寫一篇報導了。」（p. 558）這個例子闡明了爲什麼在本書中，我們一再強調，經過嚴格同儕審核的科學雜誌裡的資訊才是可靠的。這類雜誌是唯一能夠爲讀者提供科學資訊的地方。電視脫口秀沒有對資訊進行科學的審核，他們正是按麥克道格爾所說的邏輯行事：如果有人在說一件事，而且大衆對這件事感興趣，就把它搬上螢幕。測試、檢驗、眞相、證據、邏輯、論辯、資料等等，這些與脫口秀沒有任何關係，脫口秀只是提供某一時段的**娛樂**而已。要找科學資訊，請到別的地方。

媒體將科學傳播給大衆時的選擇過程，是基於以下的邏輯：言辭謹愼的科學家很少會說出什麼値得重複的名言，從而引起大衆的反響，大膽、願意冒險的人才會成爲公衆人物。當然，這也不見得總是一件壞事。舉個例子，已故天文學家和電視名人沙根（Carl Sagan），有時會比他那些保守的同事做出大膽一些的推測（Poundstone, 1999），但是他在普及天文學方面的巨大貢獻，無疑超越了他個人言論的一些不正確的瑕疵。

心理學的情形卻完全不同。多數在媒體露臉的所謂心理學家，不是沙根類型的人物，他們在自己研究領域的同行眼裡，絕對是沒有任何地位的。心理學之所以如此不同，是一系列因素綜合造成的。比起自然界其他方面，人們更迫切地希望找到有關人類自身問題的答案。比起土星環的構成、宇宙是否眞的有黑洞，人們更想知道如何減肥、心理治療是否有效、小別是否眞的更勝新婚、如何提高小孩的學業成績。這一解決自身問題的緊迫性，再加上人類問題的複雜性，致使要尋找它們的答案比較困難。這就讓我們比較淸楚地看到，爲什麼負責任的心理學家會比其他科學家更不願意在媒體上發言了。因此，在其他的學科領域，媒體的選擇過程幫他們剔除的是那些保守的科學家，取而代之的是對下結論要求不是那麼嚴謹的科學家。不幸的是，在心理學這個領域裡，媒體的選擇將所有的科學家都三振出局了！

心理學家在面對媒體記者時，爲維護科學尊嚴所把持的保守態度，往往製造了一個媒體空白，因爲不確定、有保留的說話，編造不出有趣的故事來。不管合理不合理，這個空白不久就被填補起

來了。許多自助大師和通靈騙子乘虛而入，並且在電視和電台的脫口秀節目裡吵得火熱，在大眾的心目中，這些伎倆逐漸與心理學聯結起來了。簡言之，保守主義的後果是適得其反。本來心理學家按照科學的謹慎原則向大眾發表研究結果，現在這種做法反而使得偽心理學在這個領域大行其道，從而減少了大眾和其他科學家對心理學的尊敬。

《美國心理學會會刊》（*APA Monitor*）上的一篇文章（Azar, 1998），統計了《紐約時報》（*New York Times*）在報導自然科學相關的文章時，引用科學資訊來源的比率爲51.4％，而在社會科學相關的報導中只有14.3％。無獨有偶的，心理學家西爾迪尼（Bob Cialdini, 1997）引述一項研究，發現在報紙和新聞廣播中講到自然科學研究的作者時，有80％的時候會稱他們爲科學家，而對社會科學研究的作者只有20％。相反地，這些社會科學研究的作者，有80％的時候稱爲作家或作者。這就大大削弱了社會科學研究的科學性。

有時，在報導有關「心理學」的話題時，媒體甚至懶得假裝調查過報導的科學眞實性，僅僅只是嘴上說說要注意這一問題，就算過去了。我們來看看一九九四年六月二十日《新聞週刊》的一篇文章。在花了一長欄宣傳一種新的療法之後，《新聞週刊》引述，其他新聞雜誌把這一「眼動減敏感和再加工」治療法稱爲「奇蹟」，並寫道：「不幸的是，對這一療法的熱烈反應，主要是建立在見證證據的基礎上（見第四章）。支持者還需透過科學的方法驗證，以顯示這種療法確實比其他形式的治療更具獨特優點。」（Cowley & Biddle, 1994, p. 70）那麼，你可能會問，既然還沒有足夠的科學證

據證明這一療法有效，發表這篇報導的用意何在？為了強調媒體問題給心理學造成多麼壞的影響，我們只需看看這篇文章介紹這個療法的資訊來源即可。《新聞週刊》可不是那些超市小報。你也許會詰問：**將一個未經驗證的療法發表在一個全國性週刊上，背後到底可能隱藏著什麼樣的目的？**這本週刊會被成千上萬的人讀到，其中有許多人可能正拚命地尋找可以治癒各種不同疾病的方法。在這篇文章發表兩年後，心理學家利連費爾德（Scott Lilienfeld 1996）評估了有關這一療法的療效證據，發現大多數的支持證據是來自未受控制的個案報告（見第四章）；設有控制組的實驗研究則表明沒有任何證據顯示這個療法有效。

心理學和其他學科

當然，心理學並沒壟斷對人類行為的研究。許多其他的相關學科採用不同的技術和理論視角，也對提供我們知識有所貢獻。許多關於行為的問題都要求採用交叉學科的研究方法。然而，這麼做的同時，大多數心理學家必須接受一個殘酷的事實：當這種學科整合的成果發表時，心理學家的貢獻往往被其他學科所掩蓋。

心理學家的貢獻受忽略、減低或吞占的例子不勝枚舉。例如，第一個有關電視暴力對兒童行為影響的研究，是由美國醫事總署主持的，研究結果發現二者之間存在因果關係，之後由美國醫療學

會（American Medical Association, AMA）通過一項決議，再確認該項研究結果並廣爲宣傳。當然，這一點也不令人驚訝，醫學單位主持的研究成果由醫學組織來接受認可。但這一舉措無意間造成的一個後果是，媒體的不斷將電視暴力的研究成果與美國醫療學會聯結在一起，使大衆誤以爲這個研究都是由醫學專業人士做的。事實上，絕大多數有關電視暴力對兒童行爲影響的研究，都是心理學家完成的。

心理學家的工作經常被劃入其他學科的另一個原因是，隨著時間的變遷，**心理學家**一詞的含義已經變得含混模糊了。許多心理學研究者通常會在自己的稱謂加上研究專長，例如生理心理學家、認知心理學家、工業心理學家或神經心理學家。還有人甚至用一些與**心理學**毫無關係的稱謂，例如神經科學家、認知科學家、社會生物學家、人工智慧專家和行爲研究者等。所有的這些舉動，再加上媒體認爲「心理學不是一門科學」的偏見，都導致心理學家的成就被誤劃入其他學科：生理心理學家的成果劃入生物學，認知心理學家的歸入電腦科學，工業心理學家的納入工程學和商學，等等。可悲的是，就連一位傑出的心理學研究者卡尼曼獲得諾貝爾獎，心理學的貢獻也不會被人注意到，因爲他獲得的是二〇〇二年的諾貝爾經濟學獎！當然，那是因爲諾貝爾的獎項中並沒有心理學這一項（MacCoun, 2002）。

愛默利大學（Emory University）耶克斯靈長類動物研究中心（Yerkes Primate Research Center）的負責人心理學家金恩（Frederick King, 1993）告訴我們這樣一個故事。金恩在癲癇問題的神經和行爲研究

方面經驗豐富，有一天當他花了很多時間向一位記者解釋動物模型對人類神經失調研究的重要性後，這位記者問道：「你怎麼會知道這麼多關於癲癇的資料？你只不過是一個心理學家啊！」

一九七〇年代末，法院判決了幾起有關標準測驗的訴訟案。其中的一起帕西控訴漢農（PASE vs. Hannon）案，涉及智力測驗中的文化偏見問題。審理該案的法官認爲能幫助他裁定的唯一方法是，自己親自檢查每一道測驗題，並且根據自己的直覺判斷決定這些測驗題是否眞的有文化偏見。他對自己能做出正確判決的能力深信不疑，並且對每一道測驗題寫下自己的看法（Bersoff, 1981, 1982）。這位法官斷定，在這些標準測驗中，有一組測驗中的八題和另一組中一題可能帶有偏見。這位法官沒有意識到，這樣的問題其實可以透過科學的驗證方法解答的。個人意見不僅與眞相無關，甚至可能極具誤導性。決定這些標準測驗題目是否帶有偏見，涉及複雜的統計程序和大量的資料蒐集。在爲這些評估工作進行必要的資料蒐集及發展相應的統計技術等方面，心理學家都是頗負盛名的。

諷刺的是，對這一問題所做的實際實驗研究結果發現，一般人對這些題目具有文化偏見的直覺判斷通常都是錯誤的。許多被認爲沒有偏見的題目，實際上卻含有不同的種族和社會文化歧視，而許多表面看似不公平的題目，統計結果卻沒有顯示有偏見（Sandoval & Miille, 1980）。舉個例子，韋氏成人智力量表（Wechsler Adult Intelligence Scale）在加拿大受到批評，因爲其中某次量表（資訊）的一些測驗題看起來好像偏袒美國公民，讓他們得分會比較高。例如，其中有個題目要求受試者寫出一九

五〇年之後四位美國總統的名字。於是，當這個測驗搬到加拿大實施時，題目就被「加拿大化」爲一九〇〇年之後四位加拿大總理的名字。然而，就算這樣一個非常不起眼的「常識性」改變，也產生一個小問題：加拿大公民在「總統版本」的智力測驗得分，反而比「總理版本」高。

我們是自己最可怕的敵人

爲了避免看起來在心理學的形象問題上我們好像只會責怪別人，現在是檢討心理學家自己的時候了。試圖把眞正的心理學向大衆介紹的正規心理學家，往往得不到什麼好的回報。然而，美國心理學會（APA）和美國心理協會（APS）現在都比以前更努力於促進與大衆的溝通，美國心理協會還爲了這一目的開辦一份新期刊：《大衆利益下的心理科學》（*Psychological Science in the Public Interest*）。心理學界需要在這一方面再加把勁；不然，將來如果大衆還對這一學科心存誤解，我們就只能說是自作自受了。

美國心理學會前主席福克斯（Ronald Fox, 1996）在最近的致辭中，提到心理學的一些溝通問題以及我們自己如何導致這些問題的產生：

一些經常在大眾傳媒上露臉的執業者，他們的做法是不專業的、不道德的，並且會讓

他的同行蒙羞的……我們的學科對那些不負責任、令人髮指的公開欺騙，缺乏有效的對策……在今日的世界裡，大衆經常被媒體餵食一些騙子的觀點和意見（在最近的一個電視脫口秀節目裡，有個心理學家聲稱他已經幫助許多病患回憶起上一世所受的精神創傷），而非理性的執業者的觀點。（pp. 779-780）

最後，心理學內部也存在部分反科學的態度（Coan, 1997; Watters & Ofshe, 1999）。例如，在一些心理治療圈子裡，一向拒絕對自己的療法進行科學評估。專欄作家及心理治療醫師克勞沙默（Charles Krauthammer, 1985）寫了一篇文章，論述這種態度對心理治療的誠信度所造成的嚴重損害。第一，由於沒有機制可以排除那些無效的療法，造成各種療法泛濫成災。這不僅使消費者的權益受損，還造成此一領域的混亂：「心理治療已經進入混亂的狀態，因爲……它容許太少的流派消亡，它無法在內部實現有效的淘汰，心理治療正在龍蛇混雜中走向死亡。」克勞沙默在這裡悲歎的是，不能遵循證僞原則，已經妨礙了這一領域的科學進步。

接著，克勞沙默又指出心理治療這個圈子的一個內在矛盾，一方面，他們認爲心理治療「更像一門藝術而不是科學」，因此反對採用科學的方法評估它；另一方面，他們仍然非常關注政府和健康保險公司會不會讓接受他們服務的消費者報銷。克勞沙默揭示了這兩種態度的內在矛盾：「只要心理治療業拒絕爲他們的工作提供科學證據，他們就會陷入經濟困境。畢竟，如果心理治療眞的是

一門藝術，他們應當由國家人文基金提供資助，而不是醫療機構。」與這種觀點相呼應的是，柯塔、魯傑、山德斯和霍華（S. M. Kopta, R. J. Lueger, S. M. Saunders, and K. I. Howard, 1999）在他們對心理治療效果研究的評估中寫道：「某種特殊療法的效果，必須通過實徵方法檢驗其有效性，才有足夠理由取得向保險公司和管理良好的護理公司及需要承擔責任的政府機構報銷他們服務的資格。」（p. 442）

著名的心理治療執業者彼得森（Don Peterson, 1995）也同意克勞沙默的警告。他直率地反對許多執業者所持的觀點，認爲「實徵證據總是負面的，但我的經驗告訴我不是這樣的，因此我拒絕改變我的療法。」（p. 977）他說這些人「不會從我這裡得到任何同情」（p. 977）。簡而言之，彼得森認爲心理治療執業者必須對科學證據做出回應，否則整個行業就會被大衆視爲一個「不負責任的」行業（p. 977），從而喪失了社會的支持。

在本書的早期版本中，並沒有特別強調，心理學家內部的不專業行爲和反科學態度，相當程度導致這個學科的大衆形象問題。因此有些讀者很有意見，認爲我太縱容心理學家自己的過失。在這一版，我特別加強對我們這個學科的自我批評，以求平衡這方面的論述。幸好道斯在一九九四年出版了《紙牌之屋》（*House of Cards: Psychology and Psychotherapy Built on Myth*）一書，對我完成這一平衡任務的幫助很大。如果你不相信心理學家自己是造成這個領域困境的重要原因，你就需要好好讀讀這本書。在這本需要很大勇氣才能寫成的書裡，道斯毫無保留地自曝家醜，並主張在專事研究人類

問題的心理學裡，採取科學態度對整個社會會有很大的實用價值（雖然這一潛力至今幾乎尚未開始開發）。例如，道斯寫道：「現在確實有一門眞正的心理科學，這門科學是在許多人多年來大量工作的基礎上建立起來的，但是，這門科學目前正因爲一些執業者的不專業行爲，而逐漸受到忽視、貶低和反對。這些執業者只是在嘴皮上認同這門學科而已。」(p. vii)

道斯（還有許多人，參見 Dineen, 1996; Lilienfeld, 1998; Lilienfeld et al., 2000; Mook, 2001; Watters & Ofshe, 1999）所反對的是，這個領域愈來愈關注所謂同業公會的議題——像是領發執照——而非科學研究的議題。我們可以從這一學科最古老的專業組織美國心理學會的各種活動中看出端倪。從理論上說，實施營業執照制，爲的是要保護通過特殊正統心理學訓練培養出來的專才，但是這個領域現在似乎已經忘了它的特殊性在哪裡。請你回想一下本書一開頭的討論，心理學領域的獨特性在於，它將科學研究的工具用於研究人類的行爲上；也就是說，心理學研究並不認可那些運用「直覺」能力「解讀」人的手段。確實，大量的研究文獻支持這一點。

請再回想上一章我們關於精算式預測和臨床式預測的討論。對這個問題，四十多年來的研究一致表明，在預測幾乎任何與心理學相關的事件結果時，相關變數的量化分析總比那些經驗豐富的臨床醫師更準確（Dawes et al., 1989; Faust et al., 1988; Goldberg, 1959, 1968, 1991; Meehl, 1954; Sawyer, 1966）。而且，即使是在研究心理治療的眞正（不是預測）療效這一領域，證據也表明，不管是經過培訓取得執照的心理學家，還是擁有多年豐富經驗的臨床心理學家，都不能取得比較好的療效。許多人研究

這個問題之後發現，非心理學專業的執業者（如社會工作專業人士）所取得的療效，與那些取得心理學營業執照的臨床心理學家是一樣的（Landman & Dawes, 1982; Smith et al., 1980）。事實上，即使是沒有任何資歷的助理心理專業人士所取得的療效，也與那些臨床心理學家不分伯仲（Berman & Norton, 1985; Christensen & Jacobson, 1994）。最後，沒有證據表明心理治療的臨床經驗會提升療效。一個臨床心理學家的多年工作經驗，與療效並沒有相關（Christensen & Jacobson, 1994; Landman & Dawes, 1982; Smith et al., 1980）。

如果心理學這一領域眞的想遵循依靠實徵證據來引導行動之原則，那麼它早就應該將這些研究結果公諸於世，藉以幫助美國尋找更便宜的方法以保證全民的心理健康。正如道斯（1994）所言：

> 這些結果對心理健康領域的政府公共政策有很大的影響。我們不應該浪費資源與錢財支持那些收取高價，但提供的幫助不比那些沒什麼職業訓練的好的人，他們的判斷和預測實際上比用那些依據「公開易得的」變數資料所推出的、最簡單的統計結論更糟糕。我們應當正視，心理治療的療效與職業訓練或執照無關這一事實。我們也應當正視，精心進行標準化的測驗測得的過去行為及表現的分數，才是對未來行為的最佳預測，而不是墨跡測試或由訪談中取得的印象。結論是，為了減輕病患的心理痛苦與折磨，我們應該更依賴那些聽起來還算科學的、以社區為本的心理健康專案，或依賴「助理專業人員」的幫助，這

些人能夠讓病患少花些冤枉錢。（p. 5）

心理學領域不但沒有順應這一建議，還反過來爲只准有營業執照的人執業這一行規辯護，聲稱這樣做是爲了維護心理學的科學地位，然後再用執照以掩蓋執業者內部的不科學行爲。例如，一個受過良好訓練的心理學家應當知道，我們有把握對總體的行爲做出預測，但是在預測具體的個人行爲時就有很大的不確定性（見第十章和第十一章），因此即使是最有能力的心理學家，也不應該在沒有強調這點的情況下做任何個人預測。正如道斯（1994）提到的：

> 一個心理健康專家自信地預測某人將來可能出現的行為（例如會有暴力行爲）時，已經可以肯定他不是個好的專家，因為研究證實，不論是心理健康專家或其他什麼人，都不可能有充分的證據支持他做出那麼自信的預測。（專業人士經常聲稱他們的專業角色「要求」他們做出這樣自信的判斷，而不管他們是否眞的那麼有把握。不對！他們不是被要求這樣做——他們是自願這樣做。）（p. vii）

總之，美國心理學會曾經助長心理治療領域的這股歪風。這股風氣讓人覺得，心理學家能夠透過訓練得到對個體的行爲進行「直覺式」洞察的能力。然而，實徵證據並不支持這一觀點。當被人質疑執照制度是否只是用來包庇這個行業的低能者時，學會就拿它的科學資歷作爲令箭。（針對社

會對心理學的攻擊，某位美國心理學會的主席有這樣的回應：「我們是以科學爲基礎的，這就是我們與社會工作者、諮詢師和吉普賽卜卦者之間的不同。」Dawes, 1994, p. 21）。但這個領域用來維護其科學地位的方法，卻正好揭示了他們的一個錯誤觀念：有執照的心理學家是具有獨特「臨床洞察力」的從業者。道斯的書中揭露了部分美國心理學會成員玩弄狡猾的兩面手法，某種程度上導致一九八〇年代美國心理協會的成立。這一學會的成員，是由那些對美國心理學會只關注保險組織藍十字（Blue Cross）是否可以報銷心理服務費，而不重視科學的做法感到厭倦的心理學家組成的。

利連費爾德是一位因爲事業早期對臨床心理學有貢獻而榮獲沙扣夫獎（David Shakow Award）的學者。他在頒獎典禮上不斷重申以上我們所說的論點，並警告：「在臨床心理學這一領域，我們似乎對處理僞科學這一問題完全沒有興趣，這是一個非常令人吃驚的現象，因爲這個問題的火苗已經燒到我們的後院了。」（p. 3）利連費爾德（1998）列出一九九〇年代在臨床心理學領域泛濫成災的幾種僞科學，包括：

1. 用於治療創傷的未經檢驗的奇特療法；
2. 已經證實無效的一些自閉症療法，例如溝通輔助技術（見第六章）；
3. 繼續使用一些未經充分檢驗的心理評估工具，例如各種投射測驗；
4. 各種利用潛意識做自我治療的錄音帶；

5. 使用高度暗示性的治療技巧誘導病患兒時受虐的記憶。

利連費爾德引用著名臨床研究者米爾（1993）的話：「如果我們不對這一行業進行清洗整頓，不做我們學生執著科學思考的榜樣，外行人很快就會取代我們。」（p. 728，同時參見 Mahrer, 2000）。在同一篇文章中，米爾警告我們，如果臨床心理學不採取完全科學的立場，那麼我們這些從業者也只不過是收入豐厚的算命先生而已。米爾提到這樣一種趨勢（在十一章討論過）：臨床執業者往往讓人感覺，他們對人類的行爲有「獨到」的知識。然而，實徵證據卻無法支持這一點。臨床心理學家應當更注重知識的實徵檢驗，米爾（1993）警告：「那些自以爲拿了博士學位，就能使自己免於抽樣、感知、記錄、記憶、回憶和推論上錯誤的人，是非常荒唐和自大的。」（p. 728）

心理學家培瑞茲（John Perez, 1999）主張，道斯、利連費爾德和米爾提出的觀點可以整合爲一個結論：臨床心理學家的自救運動，應當轉向謀求病患的支持，而不是那些想怎麼治療就怎麼治療的執業者。正如培瑞茲（1999）所言：「我們必須決定，我們是否想製造一個執業者想幹嘛就幹嘛的寬鬆環境，甚至犧牲科學證據把關，還是選擇保護病患的權益，以給他們最有效的治療。」（pp. 205-206）

然而，心理學領域依然被那些掛出招牌自稱爲心理學家，卻不能堅持遵循這個學科的紀律、暗地裡進行僞科學勾當的人所踐踏。當一個領域無法自律並且不能排除不科學的做法時，它就不再値

得同情了。例如，多倫多的《國家郵報》（*National Post*）曾報導加州的一位心理治療師佩里（Perry）用十分荒唐的占星術表格治療憂鬱症的故事（人們可以透過電話諮詢，諮詢費爲每小時一百五十美元）。故事的作者告訴我們：「心理學界可能很想清除像佩里之類的人，但恐怕不太容易。佩里有臨床心理學的博士學位，他的許多占星術心理學的訓練都是基於榮格的研究……這個人到今天仍在大學裡任教。」搞了半天，佩里原來是個有照的心理治療師，這個執照是由加州州政府成立的一個行爲科學審查委員會頒發的，他本人還是加州婚姻與家庭治療師協會（California Association of Marriage and Family Therapists）的成員（Milstone, 2000）。

不過，事情看似有慢慢好轉的跡象。二〇〇二年，一本新雜誌誕生了：《心理健康實踐的科學評論》（*The Scientific Review of Mental Health Practice*）（Lilienfeld, 2002）。這本雜誌致力於區分科學的治療方法與僞科學的治療，它已經得到科學心理治療實踐委員會（Council for Scientific Mental Health Practice）的認可。更令人振奮的是，至少有一些心理學組織已經痛下決心要整頓臨床實務，以及消除在實務過程中存在的那種「任憑做什麼都行」的態度。利連費爾德和洛爾（S. O. Lilienfeld and J. M. Lohr, 2000）報告了亞歷桑那州心理學資格審查委員會（Arizona Board of Psychological Examiners）吊銷一位心理學家之執照的事件。這個心理學家試圖採用僞科學的治療方法治療恐懼症，這種方法是按照預定的順序輕拍患者身體的各個部位。不必說，這種方法未經控制研究證實其有效性，亞歷桑那州這個委員會命令這位治療師停止使用這種方法，並且給他「留職查看」的處罰——一個心理學組織爲整頓內部

而嚴格禁止成員使用僞科學方法，這一事例在心理學界還是非常罕見的。

簡言之，心理學具有像化身博士（Jekyll and Hyde）那樣的雙重人格。極端嚴密的科學實驗與僞科學和反科學的態度並列而存。這一雙重人格在一九九〇年代早期關於記憶恢復—記憶錯誤的爭論中表現得淋漓盡致（Garry et al., 1999; Loftus, 1997; Loftus & Guyer, 2002; Loftus & Ketcham, 1994; Pezdek & Banks, 1996; Pezdek & Hodge, 1999; Shermer, 1997）。許多病患報告說，他們記起了幾十年前還是小孩時受虐的片段。這些片段過去一直被遺忘了，但是在治療過程中，通過干預能夠重新被喚醒。很明顯地，這些記憶中，有一部分是治療本身引發的（Campbell, 1998; Loftus & Guyer, 2002; Piper, 1998）。有些人認爲這種記憶絕對不可信，有些人則認爲它可信。在這個由富爭議性的社會虐童問題製造出來的充滿情緒性的氛圍下，心理學家提供了一些比較理性、平衡的觀點，其中較重要的是，它提供了一些客觀的科學證據（Bremner, Shobe & Kihlstrom, 2000; Clancy, Schacter, McNally, & Pitman, 2000; Pezdek & Banks, 1996）。在這裡，我們充分看到心理學的雙重人格。在經由治療干預所引發的錯誤記憶的案例中，有些是由某些不合格的、對科學無知的治療者做的，而這些治療者都是臨床心理學家。另一方面，儘管我們目前對這場爭論的結論還不是十分確定，但相關的驗證工作都是另外一些心理學家以實徵的方法艱辛努力研究的成果。心理學製造出一個問題，卻同時也解決了這個問題！

我希望這一小節的闡述，能夠幫我洗清「爲心理學家脫罪」的惡名。在本章開頭，我引用羅尼・丹佐菲的口頭禪做例子，可能令人感覺我是在維護心理學家，把心理學的形象問題歸罪於他

人。心理學家默克（2001）在他的一本研究方法書中提到我用丹佐菲的玩笑，並且評論道：「確實，通常心理學得不到應有的尊敬，但是，有時爲了某些錯誤的原因，它所得到的尊敬又完全是當之有愧的。」（p. 473）我完全同意這一感受。默克是對的，心理學的學生應當知道這個學科面臨的兩難困境。就像我在本書中表述的那樣，作爲一門研究人類行爲的科學，心理學通常沒有得到太多的尊敬。但心理學通常傳遞給大衆的形象卻是——一位臨床心理學家宣稱他有「獨到的」但不能建立在科學證據基礎上的洞察力——這時心理學卻又得到它不應該得的、過多的尊敬。這個學科的知識，通常是由那些不尊重心理學的獨特之處的人傳遞給大衆的。這一獨特之處就是，採取科學的方法驗證關於人類行爲的各種解釋及說法。

每個人不都是心理學家嗎：行爲的內隱理論

我們每個人都有關於人類行爲的一些想法或理論。沒有這些，很難想像我們怎麼活下去。從這一意義上講，我們都是心理學家。不過，是個體心理學家。區分由這種個體心理學和由科學心理學研究總結出來的知識，是十分重要的。因爲在許多大衆讀物裡，兩者的區別經常被刻意地模糊了。讓我們看看這是怎麼一回事。

與那些對行爲進行科學研究所獲得的知識相比，我們的個體心理學知識有哪些區別呢？前面已

經討論過一些。我們的個體心理學知識多數是食譜式知識。我們做某件事，是因爲我們認爲它會導致其他人做出某些相應的行爲；我們做某些事，是因爲相信這些事會幫助我們實現某些目標。這些都是所謂的食譜式知識。但是，個體心理學和科學心理學（也包括一些食譜式知識）的區分並不在於有或沒有食譜式知識而已。心理治療，在不同程度上，是在提供關於個人和環境的食譜式知識，人們可以用之解決個人的問題或讓生活更美好。其實，兩者最主要的區別是，科學心理學總是力圖透過實徵檢驗來驗證食譜式知識的有效性。

科學評估具有系統性和可控制性，這些特性個人評估並不具備、因此是有缺陷的。事實上，心理學對於決策選擇的研究表明，當人類行爲發生的情境與他們原有的信念背道而馳時，就很難在其中覺察到任何其他相關（參見 Baron, 2000）。我們只看到自己想看到的東西。雖然心理學家已經找到出現這種現象的許多原因（Kunda, 1999; Lassiter, Geers, Munhall, Ploutz-Snyder, & Breitenbecher, 2002; Stanovich, 1999，出版中），但這不是此處關注的重點。如果我們只看到自己想看到的，那麼，即使想在個人基礎上評估個體食譜式知識，那些先入爲主的偏見，也會使我們變成一個不稱職的行爲觀察者，從而使我們的評估工作更加困難。科學方法的發明，正是要避開這些單一個別觀察者的偏見。結果是，透過科學心理學產生的食譜式知識可能更精確，因爲比起個體食譜式知識，它們經過更嚴格的檢驗程序。

就像本章前面討論的，個體心理學和科學心理學之間的差別，不僅僅是對食譜式知識的檢驗。

科學總是把目標定在更高點，要在自然世界中找到比食譜式知識更多的東西。科學家想在食譜式知識之下尋求更通用、更基本的原理。但許多人對人類行爲的探究僅限於食譜式知識。事實上，對人類行爲的根本原因進行深入研究的想法，會使一些人感到害怕，他們因此總是避開這種知識。

當然，某些人的個體心理學與科學心理學有相似之處，也積極尋找更基本的心理學原理和理論。但這些人的個體理論，在許多重要方面與科學理論還是不相同。我們曾經提到，這是因爲這些個體理論無法證僞。許多人的個體心理學知識並沒有周密及合理的結構，常常只是由一堆適用於當時情境的流行語及時髦話題堆砌而成，有時還自相矛盾。它們向人們保證，不論如何都可以找到一個解釋，而且那些能夠推翻這一解釋、從而會徹底動搖人們信念的反證是不可能發生的。久之，這堆陳腔濫調就會無限上綱到可以解釋任何發生的事。正如第二章中討論的，儘管這些理論很具有撫慰的功能，但只能帶來安慰，再沒有別的了。由於它們只是在事情發生後提出的解釋，因此對將來要發生的事沒有任何預測能力。由於它們不能預測未來，也就沒有告訴我們什麼有價值的東西。心理學科的理論必須符合可證僞的標準，這就是心理科學與許多外行人的個體心理學不同之處。心理學理論能夠被證明是錯誤的，這樣知識才能不斷增長和更新。

有些個體理論不僅是一堆不可證僞的陳腔濫調，他們確實隱藏著一套心理學理論，而且是以一種比較周密而合理的結構組織起來的，同時也是有可能證僞的。然而，這類的個體心理學理論仍然存在著與個體食譜式知識一樣的一些困難。雖然它們組織周密合理並且有證僞的可能性，但仍然會

遇到評估實徵資料的問題，而這些評估有可能導致對理論的進一步修改。一個個體的觀察，並不是採用科學方法的操弄及控制建構的，它們會受制於個人的偏見，並且是以一種斷斷續續的方式取得的，因此可能導致機率推理方面的錯誤（見第十章）。在第六章中我們看到，儘管人們對移動物體有年復一年的個人體驗，他們有關自然世界中運動的直覺理論，通常都是大錯特錯的。在變化性更大的人類行爲領域中，個體理論錯誤的可能性就更大了。

抵制科學心理學的來源

正如前面討論的，千萬不要把個體心理學理論和科學心理學的知識混爲一談。這種混淆有時是蓄意製造出來的，目的是要暗中削弱心理學在大衆心目中的地位。如果「每個人都是心理學家」是指我們都有自己的內隱理論，那麼這句話是不錯的。但是它常常被隱晦地歪曲成心理學不是一門科學。

我們在第一章討論過，爲什麼建立科學心理學的想法會威脅到某些人。行爲科學知識的不斷累積，會改變那些提供心理資訊資料來源的個體、群體和組織的本質，從而發現新的心理及行爲的規律。很自然地，對那些一向以解釋及評論人類心理和行爲維生的人來說，這些新研究發現必將令他們喪失權威，因此勢必要抵制科學心理學，以確保權威不受威脅。在本書的第一章我們提到，科學

的進步會不斷地剝奪那些權威人士的地位。行星的運行、物質的本質、疾病的來源，過去曾經是神學家、哲學家和通才作家的領域；現在，天文學、物理學、醫學、遺傳學和其他學科逐漸把這些領域給拿走，放在科學家的研究範圍之內。

許多宗教都逐漸發展到不再敢聲稱，它們對宇宙結構有專門的知識。除了一些局部性的爭議之外，例如對創世說的爭論，科學與宗教之間的大戰已經成爲歷史，因爲它們已各有職司。科學家發現自然世界的結構，而許多宗教則對在應用這些發現時可能帶來的影響提出評論。宗教再也沒有與科學家一爭短長的權利了，因爲他們已經沒有能力再對科學的發現到底代表什麼意義比手劃腳了。對自然世界的裁定權，無疑已經掌握在科學家手中。

之所以會如此，是因爲信念評估的標準已經改變。很少有報紙會刊登有關土星環結構的立場鮮明的社論。爲什麼？並沒有什麼審查阻止這類社論的發表呀！很明顯的，原因是寫這類社論根本沒有用。因爲大衆相信，在這方面的知識有發言權的是科學家，而不是社論撰寫者。僅僅一百年前，報紙與那些教壇上的牧師曾對動物世界的物種起源學說大肆抨擊；現在，這類評論大都消失了。科學已經把任何理性思考者會相信這些人的評論的客觀條件徹底摧毀了。心理學則是給自然世界中的另一個巨大領域（人類行爲）帶來破壞這些條件的威脅。一百年後，標題爲「早期兒童成長是遺傳還是環境的影響？」的新聞報導，可能聽起來很愚蠢、過時，就像我們現在聽到大主教烏舍爾（Ussher）聲稱世界是在西元前四〇〇四年形成時的那種感覺。

有些人發現在面對心理學時，很難接受有一天它會到達像前述天文學（土星環結構）那樣的地步。他們頑固地堅持有權利對心理學的研究結果評頭論足，並發表對人類行爲的個人看法，即便在他們的意見與事實衝突的情況下也不罷休。當然，這裡我們不是討論權利的問題，因爲在一個自由的社會裡，每個人都有發言權，不管他說的是否正確。重要的是要意識到，許多人要的不只是發表有關人類行爲之意見的權利，他們眞正要的是**人們會相信他們這個必要條件**。談論到一些人類心理學問題時，他們希望周遭環境有利於人們接受他們的想法。這就是爲什麼認爲心理學是「任憑做什麼都行」的說法會有大量支持者，也就是心理學的說法不能由實徵方法決定對錯，它只是角度不同的問題。對於這種「任憑做什麼都行」的觀點，科學是一種威脅，因爲它有一系列嚴格的標準和程序，用以確定哪些說法可信。總之，在科學面前，這些觀點、想法都是平等的，都必須經過並通過檢驗。正是這種將錯誤理論和事實排除的能力，推動科學不斷地前進。

簡言之，許多對科學心理學的抵制來自**利益衝突**。在前面幾章中我們討論過，許多僞科學已經發展成數以百萬美元計的產業，它們利用的，就是大衆沒有意識到行爲可以用實徵方法檢驗的這一事實（在美國市場中，占星算命先生的數量是天文學家的二十倍；參見 Gilovich, 1991, p. 2）。大衆也沒有察覺支撐這些產業的技術及治療方法，如星相預測、潛意識減肥、生物節律、苦杏仁苷（laetrile，一種抗腫瘤藥）、通靈手術，經實徵檢驗後都證實是無效的。美國衆議院下的一個委員會估計過，人們每年大約花一百億美元在這些騙人的治療上，與花在正規醫療研究上的經費相比，眞是大巫見小巫

（Eisenberg et al., 1993; U.S. Congress, 1984）。

許多僞科學和虛假療法的散播者，就是在心理學領域這種「任憑做什麼都行」的氛圍中滋長的。它提供一個利用大衆無知的良好環境，因爲如果做什麼都行，大衆的消費者權益就得不到保障。正如律師哈勃（Peter Huber, 1990）所言：「在科學的邊緣和科學之外……形形色色相信順勢療法藥物、水晶和金字塔療效的人……必須誹謗正規的科學，藉以爲他們的異端邪說提供立足之地。」（p. 97）那些兜售僞科學的人總是爲了自身的利益，刻意掩蓋確實有一套科學機制可以用於檢驗行爲理論這一事實。吉瑟林（1989）警告說：「這個道理其實很簡單，人們試圖推銷他們的觀點，而眞正能評估那些觀點之好壞的人，不是購買他們產品的人。」（p. 139）在行爲理論和治療這一領域，心理學家就是那些知道如何評估產品的人。這就是爲什麼僞心理學產業極力反對科學心理學有對行爲理論進行評估的權威地位。然而，僞科學的散播者通常不需要與心理學家展開正面的鬥爭，他們採用迂迴進攻的戰略，直接走向媒體，直接向大衆推銷他們的理論或療法。大衆傳媒爲那些想要繞過科學心理學的狂徒、騙子和僞科學大開方便之門。在一九九〇年代應運而生的一大堆電視脫口秀節目裡，嘉賓並沒有被要求提供他們言論之科學證據及參考文獻，只要這些嘉賓很「有趣」，他們就可以上電視了。

世俗智慧通常包括許多妄想：人們力圖相信世界正在朝著他們期望的方向發展，而實際情況可能並非如此。爲此，科學家經常必須做些吃力不討好的工作，告訴大衆這個世界並不是像他們想像

的那樣（「不對！速食對你的健康沒有好處」）。媒體本來可以幫助科學家完成此一艱難的工作（告訴人們眞相，而不是迎合他們的期望），然而，它們卻把重心放在「娛樂」而不是事實的提供上，從而把事情搞得更糟糕。

科學確實把那些不符合最低檢驗標準、自稱爲特別知識的理論及療法清除出局。在本書裡，我們曾經簡短的探討過，在科學領域中，充分檢驗是什麼意思，不充分的檢驗又是什麼意思。內省、個人經驗和見證敘述都被認爲是不充分的檢驗。那些運用此類證據以支持其觀點的人——經常是非心理學界的評論家——會對此產生不滿，因爲他們早在當代心理學還未誕生前，就已經在使用這些不充分的證據了。

請不要以爲我想讓科學心理學承擔一個陰沈嚴肅、讓人掃興的角色。恰恰相反，比起那些整天在媒體上嘰嘰喳喳、反反覆覆的僞科學言論，科學心理學的研究發現有趣、精彩多了。同時，我們也不應認爲科學家是反對幻想、想像及欲念的——但必須在適當的情境下。梅達瓦（1990）認爲：

> 我堅信人們在適當的場合下可以隨心所欲地誇誇其談。我相信多數人在心理上需要像簡寧斯（Paul Jennings）所說的「床上冥想」（bunkrapt：你也許還記得簡寧斯打字時把「bankrupt」〔破產〕打成了「bunkrapt」）。任何人都需要幻想，我喜歡在聽華格納（Wagner）的歌劇或讀托爾金（Tolkien）的小說時幻想。但是，可不能將這個習慣帶到科學中。（p. 5）

如果我們停下來思考一分鐘，大多數人都會贊同梅達瓦的觀點。我們進入歌劇院或電影院時，經常陷入奇思妙想——這種情形卻不太可能發生在我們去看病、買保險、到託兒所給孩子註册、坐飛機或修理汽車時；這種情形也不太可能發生在我們進行心理治療、讓教育心理學家測試自己學習不佳的孩子，或把朋友帶到大學心理診所進行防止自殺諮詢的時候。心理學在追求眞相的過程中，必須像其他學科一樣，把那些毫無依據的胡思亂想、「常識」、商業廣告賣點、宗教意見、見證和妄想全數淸除。

任何一門學科一方面必須告諸社會，他們的想法和意見是有用的，另一方面，又要提醒人們區辨眞僞是一件相當困難的事。而心理學是一門正處於這種困境的新興學科。這與心理學產生的歷史時期有關，大多數學科成熟於精英控制社會結構的年代，在那個時代，普通人的意見沒有影響力；而心理學則產生於一個民主的媒體時代，忽視大眾的意見是十分危險的。許多心理學家正在努力改善這門學科與大眾在溝通方面非常糟糕的記錄，當愈來愈多的正規心理學家參與這項工作中時，勢必會增加與那些將個體心理學和科學心理學混爲一談者的衝突。

雖然每個人都有一套直覺的物理理論，但並非每個人都是物理學家。如果我們不讓這些個體的物理理論霸佔了科學物理學，就是爲物理世界的眞正科學研究做了一件好事，而這門學科的成果最終反過來會爲我們所有人做好事，因爲科學是屬於大眾的。同樣地，並非人人都是心理學家，但心理科學發現的事實和理論隨時都可以拿來應用，和豐富我們對所有人類的理解。

結語

我們現在已經到整本書的尾聲。本書雖然對如何正視心理學只提供了一幅素描，但對你理解心理學這門學科如何運作，以及如何評估新的心理學理論，有很大的幫助。我們的素描勾勒出以下幾個線索：

1. 心理學的進展是透過研究可解決的實徵問題取得的。心理學由許多不同的子領域構成，某些領域的問題比其他領域更難以研究，因此心理學的發展是不平衡的。
2. 心理學家提出可證偽的理論以解釋他們的研究發現。
3. 理論中的概念是用可操作的程序定義的，這些定義將隨著證據的累積而逐漸變化。
4. 這些理論是透過系統實徵的方法檢驗的，用這種方法蒐集的資料是公開的，也就是說，它允許其他科學家重複這些實驗並提出批評。
5. 心理學家的資料和理論，只有在那些經過同儕審核程序的科學雜誌上發表之後，才算是進入科學領域。
6. 實徵主義之所以具有系統性，是因爲它遵循眞正實驗研究的兩個特點：控制和操弄的邏輯。
7. 心理學家採用許多不同的方法取得他們的結論，這些方法的優缺點各有不同。

8. 在許多情況下，只有經過許多實驗資料的緩慢累積才能得出結論。

9. 最終發現的行爲規律，可以說總是一種機率關係。

一九六一年，英國心理學家布羅本說了一段至今仍然與當時一樣切題的話。對我們來說，它可以當作教我們如何正視心理學的一個精闢總結：

> 我們在此帶著一份質疑告終，我們不確定將來的心理學家會持有些什麼信念。事情本來就該如此。沒有人可以坐在高背椅裡就能夠抓住事實的本質。在做新的實驗之前，我們理所當然不知道結果會是怎麼樣的。我們不需要那些經常在講壇、新聞報導和學校的頒獎典禮上聽到的、關於人性的教條來引導我們。正相反，我們必須做好準備，容忍目前尚不完善的、有關行為的知識，但是堅信客觀方法的威力，總有一天會讓我們得到更完善的知識。（pp. 200-201）

參考文獻

Adler, J. E. (1998, January). Open minds and the argument from ignorance. *Skeptical Inquirer, 22*(1), 41–44.

Alcock, J. E. (1981). *Parapsychology: Science or magic?* Oxford, England: Pergamon Press.

Alcock, J. E. (1984). Parapsychology's past eight years: A lack-of-progress report. *Skeptical Inquirer, 8*, 312–320.

Alcock, J. E. (1987). Parapsychology: Science of the anomalous or search for the soul? *Behavioral and Brain Sciences, 10*, 553–643.

Alcock, J. E. (1990). *Science and supernature: A critical appraisal of parapsychology.* Buffalo, NY: Prometheus Books.

Alloy, L. B., Abramson, L. Y., & Francis, E. L. (1999). Do negative cognitive styles confer vulnerability to depression? *Current Directions in Psychological Science, 8*, 128–132.

American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: Author.

Anderson, C. A., & Anderson, K. B. (1996). Violent crime rate studies in philosophical context: A destructive testing approach to heat and Southern culture of violence effects. *Journal of Personality and Social Psychology, 70*, 740–756.

Anderson, C. A., Benjamin Jr., A. J., & Bartholow, B. D. (1998). Does the gun pull the trigger? Automatic priming effects of weapon pictures and weapon names. *Psychological Science, 9*, 308–314.

Anderson, C. A., & Dill, K. E. (2000). Video games and aggressive thoughts, feelings, and behavior in the laboratory and in life. *Journal of Personality and Social Psychology, 78*, 772–790.

Anderson, C. A., & Gunderson, K. (1991). *Paul E. Meehl: Selected philosophical and methodological papers.* Minneapolis: University of Minnesota Press.

Anderson, C. A., Lindsay, J. J., & Bushman, B. J. (1999). Research in the psychological laboratory: Truth or triviality? *Current Directions in Psychological Science, 8*, 3–9.

Anderson, D. R., Huston, A. C., Schmitt, K. L., Linebarger, D. L., & Wright, J. C. (2001). Early childhood television viewing and adolescent behavior. *Monographs of the Society for Research in Child Development, 66*(1) (Serial No. 264), 1–158.

Angell, M., & Kassirer, J. P. (1998). Alternative medicine: The risks of untested and unregulated remedies. *The New England Journal of Medicine, 339*(12), 839–841.

Arkes, H. R., & Ayton, P. (1999). The sunk cost and Concorde effects: Are humans less rational than lower animals? *Psychological Bulletin, 125*, 591–600.

Asimov, I. (1989). The relativity of wrong. *Skeptical Inquirer, 14*, 35–44.

Auletta, K. (1992). *Three blind mice: How the TV networks lost their way.* New York: Vintage Books.

Austin, E. J., & Deary, I. J. (2002). Personality dispositions. In R. J. Sternberg (Ed.), *Why smart people can be so stupid* (pp. 187–211). New Haven, CT: Yale University Press.

Azar, B. (1998, March). Are psychologists shooting themselves in the foot? *APA Monitor*, pp. 18–19.

Azar, B. (1999, November). Crowder mixes theories with humility. *APA Monitor*, p. 18.

Bachman, J. G., & Schulenberg, J. (1993). How part-time work intensity relates to drug use, problem behavior, time use, and satisfaction among high school seniors: Are these consequences or merely correlates? *Developmental Psychology, 29*, 220–235.

Bainbridge, W. S., & Stark, R. (1980). Superstitions: Old and new. *Skeptical Inquirer, 4*, 18–21.

Banaji, M. R., & Crowder, R. G. (1989). The bankruptcy of everyday memory. *American Psychologist, 44*, 1185–1193.

Baron, J. (1998). *Judgment misguided: Intuition and error in public decision making.* New York: Oxford University Press.

Baron, J. (2000). *Thinking and deciding* (3rd ed.). Cambridge, MA: Cambridge University Press.

Baron-Cohen, S. (1995). *Mindblindness: An essay on autism and theory of mind.* Cambridge, MA: The MIT Press.

Barrett, L., Dunbar, R., & Lycett, J. (2002). *Human evolutionary psychology.* Princeton, NJ: Princeton University Press.

Baumeister, R. F. (1999, January). Low self-esteem does not cause aggression. *APA Monitor,* p. 7.

Baumeister, R. F., Boden, J. M., & Smart, L. (1996). Relation of threatened egotism to violence and aggression: The dark side of high self-esteem. *Psychological Review, 103,* 5–33.

Baumeister, R. F., Bushman, B. J., & Campbell, K. W. (2000). Self-esteem, narcissism, and aggression: Does violence result from low self-esteem or from threatened egotism? *Current Directions in Psychological Science, 9,* 26–29.

Bazerman, M. (1999). *Smart money decisions.* New York: John Wiley.

Bazerman, M. (2001). Consumer research for consumers. *Journal of Consumer Research, 27,* 499–504.

Beatty, J. (1996, May). What election '96 should be about. *The Atlantic Monthly,* pp. 114–120.

Beck, A. R., & Pirovano, C. M. (1996). Facilitated communications' performance on a task of receptive language with children and youth with autism. *Journal of Autism and Developmental Disorders, 26,* 497–512.

Belsky, G., & Gilovich, T. (1999). *Why smart people make big money mistakes—And how to correct them: Lessons from the new science of behavioral economics.* New York: Simon & Schuster.

Ben-Shakhar, G., Bar-Hillel, M., Blui, Y., Ben-Abba, E., & Flug, A. (1989). Can graphological analysis predict occupational success? *Journal of Applied Psychology, 71,* 645–653.

Berkowitz, L., & Donnerstein, E. (1982). External validity is more than skin deep. *American Psychologist, 37,* 245–257.

Berliner, D. C., & Biddle, B. (1995). *The manufactured crisis: Myths, fraud, and the attack on America's public schools.* Reading, MA: Addison Wesley.

Berman, J., & Norton, N. (1985). Does professional training make a therapist more effective? *Psychological Bulletin, 98,* 401–407.

Berra, T. M. (1990). *Evolution and the myth of creation.* Stanford, CA: Stanford University Press.

Bersoff, D. (1981). Testing and the law. *American Psychologist, 36,* 1047–1056.

Bersoff, D. (1982). Larry P. and PASE: Judicial report cards on the validity of intelligence tests. In T. R. Kratochwill (Ed.), *Advances in school psychology* (pp. 145–162). Hillsdale, NJ: Erlbaum.

Best, J. (1982). Misconceptions about psychology among students who perform highly. *Psychological Reports, 51,* 239–244.

Beyerstein, B. L. (1999). Whence cometh the myth that we use only 10% of our brains? In S. Della Sala (Ed.), *Mind myths: Exploring popular assumptions about the mind and brain* (pp. 3–24). Chichester, England: John Wiley & Sons.

Birnbaum, M. H. (1999). Testing critical properties of decision making on the Internet. *Psychological Science, 10,* 399–407.

Bjorklund, D. F. (1995). *Children's thinking: Developmental function and individual differences* (2nd ed.). Pacific Grove, CA: Brooks/Cole.

Bjorklund, D. F., & Pellegrini, A. D. (2000). Child development and evolutionary psychology. *Child Development, 71,* 1687–1708.

Bjorklund, D. F., & Pellegrini, A. D. (2002). *The origins of human nature: Evolutionary developmental psychology.* Washington DC: APA.

Boden, M. A. (1990). *The creative mind: Myths and mechanisms.* New York: Basic Books.

Bok, S. (1974). The ethics of giving placebos. *Scientific American, 23,* 17–23.

Boneau, C. A. (1990). Psychological literacy: A first approximation. *American Psychologist, 45,* 891–900.

Bornstein, B. H., & Rajki, M. (1994). Extralegal factors and product liability: The influence of mock jurors' demographic characteristics and intuitions about the cause of an injury. *Behavioral Sciences and the Law, 12,* 127–147.

Bornstein, B. H., & Zickafoose, D. J. (1999). "I know I know it, I know I saw it": The stability of the confidence-accuracy rela-

tionship across domains. *Journal of Experimental Psychology: Applied, 5,* 76–88.

Bornstein, R. F. (1989). Exposure and affect. *Psychological Bulletin, 106,* 265–289.

Bower, B. (1990, July 21). The ticcing link. *Science News, 138,* 42–44.

Bower, B. (1996a, August 31). Brain images illuminate Tourette syndrome. *Science News, 150,* 133.

Bower, B. (1996b, August 24). New pitch for placebo power. *Science News, 150,* 123.

Boyer, P. J. (1999, May 17). Big guns. *The New Yorker,* pp. 54–67.

Braun, P., & Yaniv, I. (1992). A case study of expert judgment: Economists' probabilities versus base-rate model forecasts. *Journal of Behavioral Decision Making, 5,* 217–231.

Breitmeyer, B. J., & Ramey, C. T. (1986). Biological nonoptimality and quality of postnatal environment as codeterminants of intellectual development. *Child Development, 57,* 1151–1165.

Bremner, J. D., Shobe, K. K., & Kihlstrom, J. F. (2000). False memories in women with self-reported childhood sexual abuse: An empirical study. *Psychological Science, 11,* 333–337.

Brenneman, R. (1990). *Deadly blessings: Faith healing on trial.* Buffalo, NY: Prometheus Books.

Brenner, L. A., Koehler, D. J., Liberman, V., & Tversky, A. (1996). Overconfidence in probability and frequency judgments: A critical examination. *Organizational Behavior and Human Decision Processes, 65,* 212–219.

Broadbent, D. E. (1958). *Perception and communication.* New York: Pergamon Press.

Broadbent, D. (1961). *Behaviour.* New York: Basic Books.

Broadbent, D. (1973). *In defense of empirical psychology.* London: Methuen.

Bronfenbrenner, U., & Mahoney, M. (1975). The structure and verification of hypotheses. In U. Bronfenbrenner & M. Mahoney (Eds.), *Influence on human development* (pp. 131–152). Hinsdale, IL: Dryden.

Bronfenbrenner, U., McClelland, P., Wethington, E., Moen, P., & Ceci, S. J. (1996). *The state of Americans.* New York: Free Press.

Broniarczyk, S., & Alba, J. W. (1994). Theory versus data in prediction and correlation tasks. *Organizational Behavior and Human Decision Processes, 57,* 117–139.

Bronowski, J. (1956). *Science and human values.* New York: Harper & Row.

Bronowski, J. (1973). *The ascent of man.* Boston: Little, Brown.

Bronowski, J. (1974). Science, poetry, and human specificity. *American Scholar, 43,* 386–404.

Bronowski, J. (1977). *A sense of the future.* Cambridge, MA: MIT Press.

Bronowski, J. (1978a). *The common sense of science.* Cambridge, MA: Harvard University Press.

Bronowski, J. (1978b). *Magic, science, and civilization.* New York: Columbia University Press.

Budiansky, S. (1984, April 22). The meat of the matter. *Washington Post Book World,* p. 7.

Bunge, M. (1983). Speculation: Wild and sound. *New Ideas in Psychology, 1,* 3–6.

Burgess, C. A., Kirsch, I., Shane, H., Niederauer, K., Graham, S., & Bacon, A. (1998). Facilitated communication as an ideomotor response. *Psychological Science, 9,* 71–74.

Bushman, B. J., & Anderson, C. A. (2002). Violent video games and hostile expectations: A test of the general aggression model. *Personality and Social Psychology Bulletin, 28*(1679–1686).

Buss, D. M. (1985). Human mate selection. *American Scientist, 73,* 47–51.

Buss, D. M. (1999). *Evolutionary psychology: The new science of the mind.* Boston: Allyn and Bacon.

Buss, D. M. (2000). The evolution of happiness. *American Psychologist,* 55, 15–23.

Buss, D. M. (2001). Cognitive biases and emotional wisdom in the evolution of conflict between the sexes. Current Directions in *Psychological Science, 10,* 219–223.

Buss, D. M., & Barnes, M. (1986). Preferences in human mate selection. *Journal of Personality and Social Psychology, 50,* 559–570.

Byrnes, G., & Kelly, I. (1992). Crisis calls and lunar cycles: A twenty-year review. *Psychological Reports, 71,* 779–795.

Byrnes, J. P. (2001). *Minds, brains, and learning: Understanding the psychological and educational relevance of neuroscientific research.* New York: Guilford Press.

Campbell, T. (1998). *Smoke and mirrors: The devastating effect of false sexual abuse claims*. New York: Plenum.

Cardon, L. R., Smith, S., Fulker, D., Kimberling, W., Pennington, B., & DeFries, J. (1994). Quantitative trait locus for reading disability on chromosome 6. *Science, 266*, 276–279.

Carpenter, S. (1999, August 21). Kansas cuts evolution from curriculum. *Science News, 156*, 117.

Carroll, M., & Nelson, T. O. (1993). Failure to obtain a generation effect during naturalistic learning. *Memory and Cognition, 21*, 361–366.

Casscells, W., Schoenberger, A., & Graboys, T. (1978). Interpretation by physicians of clinical laboratory results. *New England Journal of Medicine, 299*, 999–1001.

Cassidy, J. (1995, October 15). Who killed the middle class? *New Yorker*, pp. 113–124.

Catrambone, R., Jones, C. M., Jonides, J., & Seifert, C. (1995). Reasoning about curvilinear motion: Using principles or analogy. *Memory and Cognition, 23*, 368–373.

Ceci, S. J., & Bruck, M. (1993a). Child witnesses: Translating research into policy. *Social Policy Report: Society for Research in Child Development, 7*(3), 1–30.

Ceci, S. J., & Bruck, M. (1993b). Suggestibility of the child witness: A historical review and synthesis. *Psychological Bulletin, 113*, 403–439.

Ceci, S. J., & Bruck, M. (1995). *Jeopardy in the courtroom*. Washington, DC: American Psychological Association.

Ceci, S. J., & Hembrooke, H. (Eds.). (1998). *Expert witnesses in child abuse cases*. Washington, DC: American Psychological Association.

Chang, E. C. (1996). Cultural differences in optimism, pessimism, and coping: Predictors of subsequent adjustment in Asian American and Caucasian American college students. *Journal of Counseling Psychology, 43*, 113–123.

Chase, A. (1977). *The legacy of Malthus*. New York: Knopf.

Cheng, C. (2001). Assessing coping flexibility in real-life and laboratory settings: A multimethod approach. *Journal of Personality and Social Psychology, 80*, 814–833.

Choi, I., Nisbett, R. E., & Norenzayan, A. (1999). Causal attribution across cultures: Variation and universality. *Psychological Bulletin, 125*, 47–63.

Christensen, A., & Jacobson, N. S. (1994). Who (or what) can do psychotherapy: The status and challenge of nonprofessional therapies. *Psychological Science, 5*, 8–14.

Christensen, D. (1999, July 17). Does practice make perfect? The benefits of busy hospitals. *Science News, 156*, 44–45.

Christensen, D. (2001, February 3). Medicinal mimicry: Sometimes, placebos work—but how? *Science News*, pp. 74–78.

Churchland, P. M. (1988). *Matter and consciousness* (rev. ed.). Cambridge, MA: MIT Press.

Churchland, P. M. (1995). *The engine of reason, the seat of the soul*. Cambridge, MA: MIT Press.

Cialdini, R. B. (1997). Professionally responsible communication with the public: Giving psychology away. *Personality and Social Psychology Bulletin, 23*, 675–683.

Clancy, S. A., Schacter, D. L., McNally, R. J., & Pitman, R. K. (2000). False recognition in women reporting recovered memories of sexual abuse. *Psychological Science, 11*, 26–31.

Clark, A. (2001). *Mindware: An introduction to the philosophy of cognitive science*. New York: Oxford University Press.

Clotfelter, C., & Cook, P. (1989). *Selling hope: State lotteries in America*. Cambridge, MA: Harvard University Press.

Clotfelter, C., & Cook, P. (1993). The "gambler's fallacy" in lottery play. *Management Science, 39*, 1521–1525.

Coan, J. A. (1997). Lost in a shopping mall: An experience with controversial research. *Ethics & Behavior, 7*, 271–284.

Coates, W., Jehle, D., & Cottington, E. (1989). Trauma and the full moon: A waning theory. *Annals of Emergency Medicine, 18*, 763–765.

Cohen, R. (1985, December 17). Despair: America's double standard. *Detroit Free Press*, p. 11.

Cole, K. C. (1998). Calculated risks. *Skeptical Inquirer, 22*(5), 32–36.

Cosmides, L., & Tooby, J. (1992). Cognitive adaptations for social exchange. In J. Barkow, L. Cosmides, & J. Tooby (Eds.),

The adapted mind (pp. 163–228). New York: Oxford University Press.
Cosmides, L., & Tooby, J. (2000). Evolutionary psychology and the emotions. In M. Lewis & J. M. Haviland-Jones (Eds.), *Handbook of emotions* (2nd ed.) (pp. 91–115). New York: Guilford Press.
Cournaud, A. (1977). The code of the scientist and its relationship to ethics. *Science, 198*, 699–705.
Cowley, G., & Biddle, N. A. (1994, June 20). Waving away the pain. *Newsweek*, pp. 70–71.
Crease, R. P., & Samios, N. P. (1991, January). Managing the unmanageable. *The Atlantic Monthly*, pp. 80–88.
Crews, F. (1993, November 18). The unknown Freud. *New York Review of Books*, pp. 55–66.
Crews, F. (1996). The verdict on Freud. *Psychological Science, 7*, 63–68.
Crews, F. (Ed.). (1998). *Unauthorized Freud: Doubters confront a legend.* New York: Viking.
Culver, R., Rotton, J., & Kelly, I. W. (1988). Moon mechanisms and myths. *Psychological Reports, 62*, 683–710.
Cummins, R. A., & Prior, M. P. (1992). Autism and assisted communication: A response to Bilken. *Harvard Educational Review, 62*, 228–241.
Curtis, B. M., & O'Keefe, J. H. (2002). Autonomic tone as a cardiovascular risk factor: The dangers of chronic fight or flight. *Mayo Clinic Proceedings, 77*(45–54).
Davidow, J., & Levinson, E. M. (1993). Heuristic principles and cognitive bias in decision making: Implications for assessment in school psychology. *Psychology in the Schools, 30*, 351–361.
Davis, D., & Holt, C. (1993). *Experimental economics.* Princeton, NJ: Princeton University Press.
Dawda, D., & Martin, J. (2001). Psychologists on psychology: The inquiry beliefs of psychologists by subdiscipline, gender, and age. *Review of General Psychology, 5*, 163–179.
Dawes, R. M. (1988). *Rational choice in an uncertain world.* San Diego: Harcourt Brace & Jovanovich.
Dawes, R. M. (1991). Probabilistic versus causal thinking. In D. Cicchetti & W. Grove (Eds.), *Thinking clearly about psychology: Essays in honor of Paul E. Meehl* (Vol. 1, pp. 235–264). Minneapolis: University of Minnesota Press.
Dawes, R. M. (1994). *House of cards: Psychology and psychotherapy built on myth.* New York: Free Press.
Dawes, R. M. (2001). *Everyday irrationality.* Boulder, CO: Westview Press.
Dawes, R. M., Faust, D., & Meehl, P. E. (1989). Clinical versus actuarial judgment. *Science, 243*, 1668–1673.
Dawkins, R. (1998). *Unweaving the rainbow.* Boston: Houghton Mifflin.
Day, S. X., & Rounds, J. (1998). Universality of vocational interest structure among racial minorities. *American Psychologist, 53*, 728–736.
Deary, I. J. (2000). *Looking down on human intelligence: From psychometrics to the brain.* Oxford: Oxford University Press.
Deary, I. J. (2001). Human intelligence differences: Towards a combined experimental-differential approach. *Trends in Cognitive Sciences, 5*(4), 164–170.
Della Sala, S. (Ed.). (1999). *Mind myths: Exploring popular assumptions about the mind and brain.* Chichester, England: John Wiley & Sons.
Dembroski, T. M., & Costa, P. T. (1987). Coronary prone behavior: Components of the Type A pattern and hostility. *Journal of Personality, 55*, 211–235.
Dembroski, T., & Costa, P. (1988). Assessment of coronary-prone behavior: A review. *Annals of Behavioral Medicine, 10*, 60–63.
Dennett, D. C. (1995). *Darwin's dangerous idea: Evolution and the meanings of life.* New York: Simon & Schuster.
Dennett, D. C. (1998). *Brainchildren: Essays on designing minds.* Cambridge, MA: MIT Press.
Dennett, D. C. (2000). Why getting it right matters: Postmodernism and truth. *Free Inquiry, 20*(1), 40–43.
Dickson, D. H., & Kelly, I. W. (1985). The "Barnum effect" in personality assessment: A review of the literature. *Psychological Reports, 57*, 367–382.
Dillon, K. (1993, Spring). Facilitated communication, autism, and ouija. *Skeptical Inquirer, 17*, 281–287.

Dineen, T. (1996). *Manufacturing victims: What the psychology industry is doing to people.* Montreal: Robert Davies.

diSessa, A. A. (1996). What do "just plain folk" know about physics? In D. R. Olson & N. Torrance (Eds.), *Handbook of education and human development: New models of learning, teaching, and schooling* (pp. 709–730). Cambridge, MA: Blackwell Publishers.

Dixon, R., & Backman, L. (Eds.). (1995). *Compensating for psychological deficits and declines: Managing losses and promoting gain.* Mahwah, NJ: Erlbaum.

Dobzhansky, T. (1973). Nothing in biology makes sense except in the light of evolution. *American Biology Teacher, 35,* 125–129.

Dodes, J. E. (1997). The mysterious placebo. *Skeptical Inquirer, 21*(1), 44–45.

Doherty, M. E., Chadwick, R., Garavan, H., Barr, D., & Mynatt, C. R. (1996). On people's understanding of the diagnostic implications of probabilistic data. *Memory and Cognition, 24,* 644–654.

Domjan, M., & Purdy, J. E. (1995). Animal research in psychology: More than meets the eye of the general psychology student. *American Psychologist, 50,* 496–503.

Dos Passos, J. (1964). *Occasions and protests.* New York: Henry Regnery.

Dowie, J., & Elstein, A. (Eds.). (1988). *Professional judgment: A reader in clinical decision making.* New York: Cambridge University Press.

Druckman, D., & Swets, J. A. (1988). *Enhancing human performance: Issues, theories and techniques.* Washington, DC: National Academy Press.

Dudley, U. (1998). Numerology: Comes the revolution. *Skeptical Inquirer, 22*(5), 29–31.

Ehrenberg, R. G., Brewer, D. J., Gamoran, A., & Williams, J. D. (2001). Class size and student achievement. *Psychological Science in the Public Interest, 2,* 1–27.

Ehri, L. C., Nunes, S., Stahl, S., & Willows, D. (2001). Systematic phonics instruction helps students learn to read: Evidence from the National Reading Panel's Meta-Analysis. *Review of Educational Research, 71,* 393–447.

Einhorn, H. J. (1986). Accepting error to make less error. *Journal of Personality Assessment, 50,* 387–395.

Eisenberg, L. (1977). The social imperatives of medical research. *Science, 198,* 1105–1110.

Eisenberg, D. M., Kessler, R., Foster, C., Norlock, F., Calkins, D., & Delbanco, T. (1993). Unconventional medicine in the United States. *The New England Journal of Medicine, 328*(4), 246–252.

Elman, J. L., Bates, E. A., Johnson, M. H., Karmiloff-Smith, A., Parisi, D., & Plunkett, K. (1996). *Rethinking innateness: A connectionist perspective on development.* Cambridge, MA: MIT Press.

Ernst, E., & Abbot, N. C. (1999). I shall please: The mysterious power of placebos. In S. Della Sala (Ed.), *Mind myths: Exploring popular assumptions about the mind and brain* (pp. 209–213). Chichester, England: John Wiley & Sons.

Estes, W. (1979). Experimental psychology: An overview. In E. Hearst (Eds.), *The first century of experimental psychology* (pp. 12–38). Hillsdale, NJ: Erlbaum.

Fackelmann, K. (1996, November 9). Gastrointestinal blues. *Science News, 150,* 302–303.

Fagley, N. (1988). Judgmental heuristics: Implications for the decision making of school psychologists. *School Psychology Review, 17,* 311–321.

Falk, R. (1989). Judgment of coincidences: Mine versus yours. *American Journal of Psychology, 102,* 477–493.

Falk, R., & MacGregor, D. (1983). The surprisingness of coincidences. In P. Humphreys, O. Svenson, & A. Vari (Eds.), *Analysis and aiding decision processes* (pp. 489–502). Budapest: Akademiai Kiado.

Fancher, R. (1985). *The intelligence men.* New York: W. W. Norton.

Fantino, E., & Esfandiari, A. (2002). Probability matching: Encouraging optimal responding in humans. *Canadian Journal of Experimental Psychology, 56,* 58–63.

Farber, I. B., & Churchland, P. S. (1995). Consciousness and the neurosciences: Philosophical and theoretical issues. In M. S. Gazzaniga (Ed.), *The cognitive neurosciences* (pp. 1295–1306). Cambridge, MA: MIT Press.

Faust, D., Hart, T., Guilmette, T., & Arkes, H. (1988). Neuropsychologists' capacity to detect adolescent malingerers. *Profes-*

sional Psychology: Research and Practice, 19, 578–582.

Feeney, D. (1987). Human rights and animal welfare. *American Psychologist, 42*, 593–599.

Fernald, D. (1984). *The Hans legacy*. Hillsdale, NJ: Erlbaum.

Ferrari, M., & Sternberg, R. J. (1998). The development of mental abilities and styles. In D. Kuhn & R. S. Siegler (Eds.), *Handbook of child psychology* (Vol. 2, pp. 899–946). New York: John Wiley.

Finkel, A. M. (1996, May). Who's exaggerating? *Discover, 17*(5), 48–54.

Finn, J. D., & Achilles, C. M. (1999). Tennessee's class size study: Findings, implications, misconceptions. *Educational Evaluation and Policy Analysis, 21*, 97–109.

Foster, E. A., Jobling, M. A., Taylor, P. G., Donnelly, P., Deknijff, P., Renemieremet, J., Zerjal, T., & Tyler-Smith, C. (1998). Jefferson fathered slave's last child. *Nature, 396*, 27–28.

Foster, K. R., & Huber, P. W. (1999). *Judging science: Scientific knowledge and the federal courts*. Cambridge, MA: MIT Press.

Fox, R. E. (1996). Charlatanism, scientism, and psychology's social contract. *American Psychologist, 51*, 777–784.

Frank, R. H. (1999). *Luxury fever: Why money fails to satisfy in an era of excess*. New York: Free Press.

Frank, R. H., & Cook, P. J. (1995). *The winner-take-all society*. New York: Free Press.

Fridson, M. S. (1993). *Investment illusions*. New York: John Wiley.

Fried, S. B. (1994). *American popular psychology: An interdisciplinary research guide*. New York: Garland Publishing.

Fried, S. B. (1998). An undergraduate course in American popular psychology. *Teaching of Psychology, 25*, 38–40.

Fried, S. B., & Schultis, G. A. (1995). *The best self-help and self awareness books*. Chicago: American Library Association.

Friedman, M., & Booth-Kewley, S. (1987). The "disease-prone personality." *American Psychologist, 41*, 539–555.

Friedman, M., & Ulmer, D. (1984). *Treating Type A behavior and your heart*. New York: Knopf.

Friedrich, J., Buday, E., & Kerr, D. (2000). Statistical training in psychology: A national survey and commentary on undergraduate programs. *Teaching of Psychology, 27*, 248–257.

Friedrich-Cofer, L., & Huston, A. (1986). Television violence and aggression: The debate continues. *Psychological Bulletin, 100*, 364–371.

Frith, U. (1989). *Autism: Explaining the enigma*. Oxford, England: Basil Blackwell.

Frith, U. (1993, June). Autism. *Scientific American, 268*(6), 108–114.

Gage, N. L. (1996). Confronting counsels of despair for the behavioral sciences. *Educational Researcher, 25*(3), 5–15.

Gage, N. L., & Berliner, D. (1984). *Educational psychology* (3rd ed.). San Francisco: Freeman.

Gal, I., & Baron, J. (1996). Understanding repeated simple choices. *Thinking and Reasoning, 2*, 81–98.

Galaburda, A. (1994). Developmental dyslexia and animal studies: At the interface between cognition and neurology. *Cognition, 50*, 133–149.

Garb, H. N., Florio, C. M., & Grove, W. M. (1998). The validity of the Rorschach and the Minnesota Multiphasic Personality Inventory. *Psychological Science, 9*, 402–404.

Garb, H. N., & Schramke, C. J. (1996). Judgment research and neuropsychological assessment: A narrative review and meta-analysis. *Psychological Bulletin, 120*, 140–153.

Gardner, M. (1972, April). Mathematical games: Why the long arm of coincidence is usually not as long as it seems. *Scientific American, 227*(4), 110–112.

Gardner, M. (2001). Facilitated communication: A cruel farce. *Skeptical Inquirer, 25*(1), 17–19.

Garry, M., Frame, S., & Loftus, E. F. (1999). Lie down and tell me about your childhood. In S. Della Sala (Ed.), *Mind myths: Exploring popular assumptions about the mind and brain* (pp. 113–124). Chichester, England: John Wiley & Sons.

Gawande, A. (1998, February 8). No mistake. *The New Yorker*, pp. 74–81.

Gawande, A. (1999, February 9). The cancer-cluster myth. *The New Yorker*, pp. 35–37.

Gazzaniga, M. S. (2000). *The new cognitive neurosciences* (2nd ed.). Cambridge, MA: MIT Press.

Geschwind, N. (1985). The biology of dyslexia. In D. Gray & J. Kavanagh (Eds.), *Behavioral measures of dyslexia* (pp. 19–37). Parkton, MD: York Press.

Ghiselin, M. T. (1989). *Intellectual compromise.* New York: Paragon House.

Gibson, E. J. (1994). Has psychology a future? *Psychological Science, 5,* 69–76.

Gigerenzer, G. (1984). External validity of laboratory experiments: The frequency-validity relationship. *American Journal of Psychology, 97,* 185–195.

Gigerenzer, G. (2002). *Calculated risks: How to know when numbers deceive you.* New York: Simon & Schuster.

Gilovich, T. (1991). *How we know what isn't so: The fallibility of human reason in everyday life.* New York: Free Press.

Gilovich, T., Griffin, D., & Kahneman, D. (Eds.). (2002). *Heuristics and biases: The psychology of intuitive judgment.* New York: Cambridge University Press.

Gilovich, T., Vallone, R., & Tversky, A. (1985). The hot hand in basketball: On the misperception of random sequences. *Cognitive Psychology, 17,* 295–314.

Gladwell, M. (1996, July 8). Conquering the coma. *The New Yorker,* pp. 34–40.

Glassner, B. (1999). *The culture of fear: Why Americans are afraid of the wrong things.* New York: Basic Books.

Gleitman, H. (1981). *Psychology.* New York: W. W. Norton.

Goldberg, L. R. (1959). The effectiveness of clinicians' judgments: The diagnosis of organic brain damage from the Bender Gestalt Test. *Journal of Consulting Psychology, 23,* 25–33.

Goldberg, L. R. (1968). Simple models or simple processes? Some research on clinical judgments. *American Psychologist, 23,* 483–496.

Goldberg, L. R. (1991). Human mind versus regression equation: Five contrasts. In D. Cicchetti & W. Grove (Eds.), *Thinking clearly about psychology: Essays in honor of Paul E. Meehl* (Vol. 1, pp. 173–184). Minneapolis: University of Minnesota Press.

Goldstein, M., & Goldstein, I. (1978). *How we know: An exploration of the scientific process.* New York: Plenum Press.

Goldstein, W. M., & Hogarth, R. M. (Eds.). (1997). *Research on judgment and decision making.* Cambridge, England: Cambridge University Press.

Goswami, U. (1998). *Cognition in children.* Hove, England: Psychology Press.

Gould, S. J. (1987). Justice Scalia's misunderstanding. *Natural History, 96,* 14–21.

Greenberger, D., & Steinberg, L. (1986). *When teenagers work.* New York: Basic Books.

Greenwald, A. G., Spangenberg, E., Pratkanis, A., & Eskenazi, J. (1991). Double-blind tests of subliminal self-help audiotapes. *Psychological Science, 2,* 119–122.

Grice, G. (2001, July 30). Slice of life: How a convicted killer's corpse brought anatomy into the digital age. *The New Yorker,* pp. 36–41.

Griffin, D., & Tversky, A. (1992). The weighing of evidence and the determinants of confidence. *Cognitive Psychology, 24,* 411–435.

Griggs, R. A., Proctor, D. L., & Bujak-Johnson, A. (2002). The nonexistent common core. *American Psychologist, 57,* 452–453.

Grimmer, M. R. (1992). Searching for security in the mystical: The function of paranormal beliefs. *Skeptical Inquirer, 16,* 173–176.

Groopman, J. (1999, May 10). Pet scan. *The New Yorker,* pp. 46–50.

Hacking, I. (1975). *The emergence of probability.* Cambridge, England: Cambridge University Press.

Hacohen, M. C. (2000). *Karl Popper: The formative years, 1902–1945.* Cambridge, England: Cambridge University Press.

Hagen, M. A. (2001). Damaged goods? What, if anything, does science tell us about the long-term effects of childhood sexual abuse? *Skeptical Inquirer, 25*(1), 54–59.

Haggbloom, S. J., Warnick, R., Warnick, J. E., Jones, V. K., Yarbrough, G. L., Russell, T. M., Borecky, C. M., McGahhey, R., Powell, J. L., Beavers, J., & Monte, E. (2002). The 100 most eminent psychologists of the 20th century. *Review of General Psychology, 6,* 139–152.

Halpern, D. F. (1998). Teaching critical thinking for transfer across domains: Dispositions, skills, training, and metacognitive monitoring. *American Psychologist, 53,* 449–455.

Hammond, K. R. (1996). *Human judgement and social policy.* New York: Oxford University Press.

Hammond, K. R., Harvey, L. O., & Hastie, R. (1992). Making better use of scientific knowledge: Separating truth from justice. *Psychological Science, 3,* 80–87.
Harlow, H. F. (1958). The nature of love. *American Psychologist, 13,* 673–685.
Harlow, H. F., & Suomi, S. J. (1970). The nature of love—Simplified. *American Psychologist, 25,* 161–168.
Harrington, A. (1997). *The placebo effect.* Cambridge, MA: Harvard University Press.
Hastie, R., & Dawes, R. M. (2001). *Rational choice in an uncertain world.* Thousand Oaks, CA: Sage.
Hearst, E. (1979). *The first century of experimental psychology.* Hillsdale, NJ: Erlbaum.
Heath, L., Tindale, R., Edwards, J., Posavac, E., Bryant, F., Henderson-King, E., Suarez-Balcazar, Y., & Myers, J. (Eds.). (1994). *Applications of heuristics and biases to social issues.* New York: Plenum Press.
Heinsman, D. T., & Shadish, W. (1996). Assignment methods in experimentation: When do nonrandomized experiments approximate answers from randomized experiments? *Psychological Methods, 1,* 154–169.
Higbee, K. L., & Clay, S. L. (1998). College students' beliefs in the ten-percent myth. *The Journal of Psychology, 132,* 469–476.
Hines, T. M. (1988). *Pseudoscience and the paranormal.* Buffalo, NY: Prometheus Books.
Hines, T. M. (1998). Comprehensive review of biorhythm theory. *Psychological Reports, 83,* 19–64.
Hollon, S. D., Thase, M. E., & Markowitz, J. C. (2002). Treatment and prevention of depression. *Psychological Science in the Public Interest, 3,* 39–77.
Holton, G. (1996). *Einstein, history, and other passions: The rebellion against science at the end of the twentieth century.* Reading, MA: Addison-Wesley.
Holton, G., & Roller, D. (1958). *Foundations of modern physical science.* Reading, MA: Addison Wesley.
Hooft, G. T. (2000). Physics and the paranormal. *Skeptical Inquirer, 24*(2), 27–30.
Huber, P. W. (1990). *Pathological science in court.* Daedalus, 119, 97–118.
Hudson, A., Melita, B., & Arnold, N. (1993). A case study assessing the validity of facilitated communication. *Journal of Autism and Developmental Disorders, 23,* 165–173.
Huesmann, L. R., & Eron, L. D. (1986). *Television and the aggressive child: A cross-national comparison.* Hillsdale, NJ: Erlbaum.
Humphrey, N. (1996). *Leaps of faith.* New York: Copernicus.
Hunt, M. (1999). *The new know-nothings: The political foes of the scientific study of human nature.* New Brunswick, NJ: Transaction Publishers.
Hyman, R. (1981). Cold reading: How to convince strangers that you know all about them. In K. Frazier (Eds.), *Paranormal borderlands of science* (pp. 79–96). Buffalo, NY: Prometheus Books.
Hyman, R. (1992). What does Goodfellow's classic analysis tell researchers today? *Journal of Experimental Psychology: General, 121,* 128–129.
Hyman, R. (1996). Evaluation of the military's twenty-year program on psychic spying. *Skeptical Inquirer, 20*(2), 21–26.
Hynd, G. W., Clinton, A., & Hiemenz, J. R. (1999). The neuropsychological basis of learning disabilities. In R. J. Sternberg & L. Spear-Swerling (Eds.), *Perspectives on learning disabilities* (pp. 60–79). New York: Westview/HarperCollins.
If it sounds just ducky, it probably quacks. (1990, March 23). *Detroit Free Press.*
Immen, W. (1996, August 8). Could you repeat that in Klingon? *Globe & Mail (Toronto).*
Jacobson, J. W., Mulick, J. A., & Schwartz, A. A. (1995). A history of facilitated communication: Science, pseudoscience, and antiscience. *American Psychologist, 50,* 750–765.
Jacobson, J. W., Mulick, J. A., & Schwartz, A. A. (1996). If a tree falls in the woods. *American Psychologist, 51,* 988–989.
Johnson, H. (1991). *Sleepwalking through history: America in the Reagan years.* New York: Anchor Books.
Kahne, J. (1996). The politics of self-esteem. *American Educational Research Journal, 33,* 3–22.
Kahneman, D. (1973). *Attention and effort.* Englewood Cliffs, NJ: Prentice Hall.
Kahneman, D., & Frederick, S. (2002). Representativeness revisited: Attribute substitution in intuitive judgment. In T.

Gilovich, D. Griffin, & D. Kahneman (Eds.), *Heuristics and biases: The psychology of intuitive judgment* (pp. 49–81). New York: Cambridge University Press.

Kahneman, D., & Tversky, A. (Eds.). (2000). *Choices, values, and frames*. Cambridge, England: Cambridge University Press.

Kalat, J. W. (1995). *Biological psychology* (5th ed.). Pacific Grove, CA: Brooks/Cole.

Kaminer, W. (1999). *Sleeping with extra-terrestrials: The rise of irrationalism and the perils of piety*. New York: Pantheon.

Kassin, S. M., Tubb, V. A., Hosch, H. M., & Memon, A. (2001). On the "general acceptance" of eyewitness testimony research. *American Psychologist, 56*, 405–416.

Kelly, I. W. (1997). Modern astrology: A critique. *Psychological Reports, 81*, 931–962.

Kelly, I. W. (1998). Why astrology doesn't work. *Psychological Reports, 82*, 527–546.

Kelly, I., Rotton, J., & Culver, R. (1985). The moon was full and nothing happened: A review of studies on the moon and human behavior and lunar beliefs. *Skeptical Inquirer, 10*, 129–143.

Kenrick, D. T. (2001). Evolutionary psychology, cognitive science, and dynamical systems: Building an integrative paradigm. *Current Directions in Psychological Science, 10*(1), 13–17.

Kershaw, A. (1991, July 21). *The psychic guide to hiring and firing*. London Independent.

Kety, S. (1974). Preface. In J. Swazey, *Chlorpromazine in psychiatry: A study of therapeutic intervention*. Cambridge, MA: MIT Press.

Kim, J. (1994, September). Watch out for investing newsletters luring you with outdated returns. *Money Magazine*, pp. 12–13.

Kimble, G. A. (1999). Functional behaviorism: A plan for unity in psychology. *American Psychologist, 54*, 981–988.

King, F. (1993, March). Science literacy and public support of science. *American Psychological Society Observer, 6*(2), 2–11.

King, R. N., & Koehler, D. J. (2000). Illusory correlations in graphological inference. Journal of Experimental Psychology: *Applied, 6*, 336–348.

Kirsch, I. (Ed.). (1999). *How expectancies shape experience*. Washington, DC: American Psychological Association.

Knight, G. P., Johnson, L., Carlo, G., & Eisenberg, N. (1994). A multiplicative model of the dispositional antecedents of a prosocial behavior: Predicting more of the people more of the time. *Journal of Personality and Social Psychology, 66*, 178–183.

Koehler, J. J. (1993). The normative status of base rates at trial. In N. J. Castellan (Eds.), *Individual and group decision making* (pp. 137–149). Hillsdale, NJ: Erlbaum.

Kohn, A. (1990). *You know what they say . . . : The truth about popular beliefs*. New York: HarperCollins.

Kolb, B., & Whishaw, I. (1990). *Fundamentals of human neuropsychology* (3rd ed.). New York: Freeman.

Kopta, S. M., Lueger, R. J., Saunders, S. M., & Howard, K. I. (1999). Individual psychotherapy outcome and process research. *Annual Review of Psychology, 50*, 441–469.

Kramer, S. P. (1987). *How to think like a scientist: Answering questions by the scientific method*. New York: Thomas Y. Crowell Junior Books.

Krauthammer, C. (1985, December 29). The twilight of psychotherapy? *Detroit News*.

Krippner, S. (1977). *Advances in parapsychological research: 1. Psychokinesis*. New York: Plenum Press.

Krull, D. S., Hui-Min Loy, M., Lin, J.,Wang, C., Chen, S., & Zhao, X. (1999). The fundamental attribution error: Correspondence bias in individualist and collectivist cultures. *Personality and Social Psychology Bulletin, 25*, 1208–1219.

Kuhn, D., Weinstock, M., & Flaton, R. (1994). How well do jurors reason? Competence dimensions of individual variation in a juror reasoning task. *Psychological Science, 5*, 289–296.

Kunda, Z. (1999). *Social cognition: Making sense of people*. Cambridge, MA: MIT Press.

Kushner, H. I. (1999). *A cursing brain? The histories of Tourette Syndrome*. Cambridge, MA: Harvard University Press.

Landman, J. T., & Dawes, R. M., (1982). Psychotherapy Outcome. *American Psychologist, 31*, 504–516.

Langer, E. J. (1975). The illusion of control. *Journal of Personality and Social Psychology, 32*, 311–328.

Langewiesche, W. (1993, December). The turn. *The Atlantic Monthly*, pp. 115–122.

Lassiter, G. D., Geers, A. L., Munhall, P. J., Ploutz-Snyder, R. J., & Breitenbecher, D. L. (2002). Illusory causation: Why it occurs. *Psychological Science, 13*, 299–305.

Latane, B., & Darley, J. (1970). *The unresponsive bystander: Why doesn't he help?* New York: Appleton-Century-Crofts.

Lazar, I., Darlington, R., Murray, H., Royce, J., & Sniper, A. (1982). Lasting effects of early education: A report from the Consortium of Longitudinal Studies. *Monographs of the Society for Research in Child Development, 47* (Serial No. 195).

Leary, D. E. (2001). One big idea, one ultimate concern. *American Psychologist, 56*, 425–432.

Leavy, J. (1992). Our spooky presidential coincidences contest. *Skeptical Inquirer, 16*, 316–319.

Lee, V., Brooks-Gunn, J., Schnur, E., & Liaw, F. (1990). Are Head Start's effects sustained? A longitudinal follow-up comparison of disadvantaged children attending Head Start, no preschool, and other preschool programs. *Child Development, 61*, 495–507.

Lees-Haley, P. R. (1997). Manipulation of perception in mass tort litigation. *Natural Resources & Environment, 12*, 64–68.

Leibowitz, H. W. (1996). The symbiosis between basic and applied research. *American Psychologist, 51*, 366–370.

Leli, D., & Filskov, S. (1981). Clinical-actuarial detection and description of brain impairment with the W-B Form 1. *Journal of Clinical Psychology, 37*, 623–629.

Lerner, M. J., & Miller, D. T. (1978). Just world research and the attribution process: Looking back and ahead. *Psychological Bulletin, 85*, 1030–1051.

Leslie, A. M. (2000). "Theory of mind" as a mechanism of selective attention. In M. S. Gazzaniga (Ed.), *The new cognitive neurosciences* (2nd ed.) (pp. 1235–1247). Cambridge, MA: MIT Press.

Levin, J. R., & O'Donnell, A. M. (2000). What to do about educational research's credibility gaps? *Issues in Education: Contributions from Educational Psychology, 5*, 1–87.

Levine, K., Shane, H., & Wharton, R. (1994). What if . . . : A plea to professionals to consider the risk-benefit ratio of facilitated communication. *Mental Retardation, 32*, 300–304.

Lewis, M. (1997). *Trail fever.* New York: Knopf.

Li, C. (1975). *Path analysis: A primer.* Pacific Grove, CA: Boxwood Press.

Lichtenstein, S., Slovic, P., Fischhoff, B., Layman, M., & Combs, B. (1978). Judged frequency of lethal events. Journal of Experimental Psychology: *Human Learning and Memory, 4*, 551–578.

Lilienfeld, S. O. (1996). EMDR treatment: Less than meets the eye? *Skeptical Inquirer, 20*(1), 25–31.

Lilienfeld, S. O. (1998). Pseudoscience in contemporary clinical psychology: What it is and what we can do about it. *The Clinical Psychologist, 51*(4), 3–9.

Lilienfeld, S. O. (1999). Projective measures of personality and psychopathology: How well do they work? *Skeptical Inquirer, 23*(5), 32–39.

Lilienfeld, S. O. (2002). New journal dedicated to distinguishing science from pseudoscience in mental health. *Skeptical Inquirer, 26*(3), 7–8.

Lilienfeld, S. O., & Lohr, J. M. (2000). Thought field therapy practitioners and educators sanctioned. *Skeptical Inquirer, 24*(2), 5.

Lilienfeld, S. O., Lohr, J. M., & Moirer, D. (2001). The teaching of courses in the science and pseudoscience of psychology: Useful resources. *Teaching of Psychology, 28*(3), 182–191.

Lilienfeld, S. O.,Wood, J. M., & Garb, H. N. (2000). The scientific status of projective techniques. *Psychological Science in the Public Interest, 1*, 27–66.

Lipsey, M. W., & Wilson, D. B. (1993). The efficacy of psychological, educational, and behavioral treatment: Confirmation from meta-analysis. *American Psychologist, 48*, 1181–1209.

Loftus, E. F. (1993). Psychologists in the eyewitness world. *American Psychologist, 48*, 550–552.

Loftus, E. F. (1994). The repressed memory controversy. *American Psychologist, 49*, 443–445.

Loftus, E. F. (1997, September). Creating false memories. *Scientific American, 276*(9), 70–75.

Loftus, E. F., & Guyer, M. J. (2002, May/June). Who abused Jane Doe: The hazards of the single case history. *Skeptical Inquirer, 26*(3), 24–32.

Loftus, E. F., & Ketcham, K. (1994). *The myth of repressed memory: False memories and allegations of sexual abuse.* New York: St. Martin's Press.

Lopes, L., & Oden, G. (1987). Distinguishing between random and nonrandom events. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13,* 392–400.

MacCoun, R. (2002, December). Why a psychologist won the Nobel Prize in economics. *APS Observer, 15*(10), 1–8.

MacDonald, A. (1990, October 3). Parents fear wrong things, survey suggests. *Ann Arbor News* (taken from The New York Times wire service).

MacDougall, C. D. (1983). *Superstition and the press.* Buffalo, NY: Prometheus Books.

Macmillan, M. (1997). *Freud evaluated.* Cambridge, MA: MIT Press.

Magee, B. (1985). *Philosophy and the real world: An introduction to Karl Popper.* LaSalle, IL: Open Court.

Mahrer, A. R. (2000). Philosophy of science and the foundations of psychotherapy. *American Psychologist, 55,* 1117–1125.

Malkiel, B. G. (1999). *A random walk down Wall Street.* New York: W. W. Norton.

Margolis, H. (1996). *Dealing with risk.* Chicago: University of Chicago Press.

Marinoff, L. (1999). *Plato, not Prozac: Applying philosophy to everyday problems.* New York: HarperCollins.

Market Facts, I. (1984). *1983 consumer research study on reading and book purchasing.* New York: Book Industry Study Group.

Marks, D. F. (2001). *The psychology of the psychic.* Buffalo, NY: Prometheus Books.

Marks, D., & Kammann, R. (1980). *The psychology of the psychic.* Buffalo, NY: Prometheus Books.

Marshall, E. (1980). Police science and psychics. *Science, 210,* 994–995.

Martin, B. (1998). Coincidences: Remarkable or random? *Skeptical Inquirer, 22*(5), 23–28.

Maynard Smith, J. (1998). *Evolutionary genetics* (2nd ed.). Oxford: Oxford University Press.

McBride-Chang, C., & Kail, R. V. (2002). Cross-cultural similarities in the predictors of reading acquisition. *Child Development, 73,* 1392–1407.

McCloskey, M. (1983, April). Intuitive physics. *Scientific American, 248*(4), 122–130.

McCorduck, P. (1979). *Machines who think.* San Francisco: Freeman.

McCullough, M. L. (2001). Freud's seduction theory and its rehabilitation: A saga of one mistake after another. *Review of General Psychology, 5,* 3–22.

McCutcheon, L. E., Furnham, A., & Davis, G. (1993). A cross-national comparison of students' misconceptions about psychology. *Psychological Reports, 72,* 243–247.

McFall, R. M., & Treat, T. A. (1999). Quantifying the information value of clinical assessments with signal detection theory. *Annual Review of Psychology, 50,* 215–241.

McGraw, K. O.,Tew, M. D., & Williams, J. E. (2000). The integrity of web-delivered experiments: Can you trust the data? *Psychological Science, 11,* 502–506.

Medawar, P. B. (1967). *The art of the soluble.* London: Methuen.

Medawar, P. B. (1979). *Advice to a young scientist.* New York: Harper & Row.

Medawar, P. B. (1984). *The limits of science.* New York: Harper & Row.

Medawar, P. B. (1990). *The threat and the glory.* New York: HarperCollins.

Medawar, P. B., & Medawar, J. S. (1983). *Aristotle to zoos: A philosophical dictionary of biology.* Cambridge, MA: Harvard University Press.

Meehl, P. E. (1954). *Clinical versus statistical prediction: A theoretical analysis and review of the literature.* Minneapolis: University of Minnesota Press.

Meehl, P. E. (1986). Causes and effects of my disturbing little book. *Journal of Personality Assessment, 50,* 370–375.

Meehl, P. E. (1991). Law and the fireside inductions: Some reflections of a clinical psychologist. In C. Anderson & K. Gunderson (Eds.), *Paul E. Meehl: Selected philosophical and methodological papers* (pp. 440–480). Minneapolis: University of Minnesota Press.

Meehl, P. E. (1993). Philosophy of science: Help or hindrance? *Psychological Reports, 72,* 707–733.

Menon, T., Morris, M. W., Chiu, C., & Hong, Y. (1999). Culture and the construal of agency: Attribution to individual versus group dispositions. *Journal of Personality and Social Psychology, 76*, 701–717.

Metalsky, G. I., & Joiner, T. E. (1992). Vulnerability to depressive symptomatology: A prospective test of the diathesis-stress and causal mediation components of the hopelessness theory of depression. *Journal of Personality and Social Psychology, 63*, 667–675.

Miller, M., Azrael, D., & Hemmingway, D. (2002). Firearm availability and unintentional firearm deaths, suicide, and homicide among 5–14 year olds. *Journal of Trauma, Injury, Infection, and Critical Care, 52*, 267–275.

Miller, N. E. (1985a). Rx: Biofeedback. *Psychology Today, 19*, 54–59.

Miller, N. E. (1985b). The value of behavioral research on animals. *American Psychologist, 40*, 423–440.

Miller, T. Q., Turner, C., Tindale, R., Posavac, E., & Dugoni, B. (1991). Reasons for the trend toward null findings in research on Type A behavior. *Psychological Bulletin, 110*, 469–485.

Milstone, C. (2000, February 28). California psychotherapist uses astrological charts to treat stress and depression. *National Post (Toronto)*, p. D3.

Milton, J., & Wiseman, R. (1999). Does psi exist? Lack of replication of an anomalous process of information transfer. *Psychological Bulletin, 125*, 387–391.

Mishel, L. (1995, Fall). Rising tides, sinking wages. *American Prospect*, pp. 60–64.

Monk, R. (1990). *Ludwig Wittgenstein: The duty of genius.* New York: Free Press.

Mook, D. G. (1982). *Psychological research: Strategies and tactics.* New York: Harper & Row.

Mook, D. G. (1983). In defense of external invalidity. *American Psychologist, 38*, 379–387.

Mook, D. G. (1989). The myth of external validity. In L. W. Poon, D. C. Rubin, & B. A. Wilson (Eds.), *Everyday cognition in adulthood and late life* (pp. 25–43). Cambridge, England: Cambridge University Press.

Mook, D. G. (2001). *Psychological research: The ideas behind the methods.* New York: Norton.

Moore, T. E. (1995). Subliminal self-help auditory tapes: An empirical test of perceptual consequences. *Canadian Journal of Behavioural Science, 27*, 9–20.

Moore, T. E. (1996). Scientific consensus and expert testimony: Lessons from the Judas Priest trial. *Skeptical Inquirer, 20*(6), 32–38.

Mulick, J., Jacobson, J., & Kobe, F. (1993, Spring). Anguished silence and helping hands: Autism and facilitated communication. *Skeptical Inquirer, 17*, 270–280.

Murchison, C. (1934). *Handbook of general experimental psychology.* Worcester, MA: Clark University Press.

Murray, D., Schwartz, J., & Lichter, S. R. (2001). *It ain't necessarily so: How media make and unmake the scientific picture of reality.* Lanham, MD: Rowman & Littlefield.

Murstein, B. (1980). Mate selection in the 1970s. *Journal of Marriage and Family, 42*, 51–66.

Nathan, P. E., & Gorman, J. M. (Ed.). (1998). *A guide to treatments that work.* New York: Oxford University Press.

National Safety Council. (1990). *Accident facts* (1990 edition). Chicago: Author.

National Safety Council (2001). *Report on Injuries in America, 2001.* Retrieved March 29, 2002, from www.nsc.org/library/rept2000.htm.

National Science Board (2001). Science indicators 2000: Belief in the paranormal or pseudoscience. *Skeptical Inquirer, 25*(1), 12–15.

Neter, E., & Ben-Shakhar, G. (1989). Predictive validity of graphological inferences: A meta-analysis. *Personality and Individual Differences, 10*, 737–745.

Nickerson, R. (1992). *Looking ahead: Human factors challenges in a changing world.* Hillsdale, NJ: Erlbaum.

Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology, 2*, 175–220.

Nickerson, R. S. (1999). Basic versus applied research. In R. J. Sternberg (Ed.), *The nature of cognition* (pp. 409–423). Cambridge, MA: MIT Press.

Nickerson, R. S. (2002). The production and perception of randomness. *Psychological Review, 109*, 330–357.

Niemeyer, W., & Starlinger, I. (1981). Do the blind hear better? *Audiology, 20*, 503–515.

Nisbett, R. E., Peng, K., Choi, I., & Norenzayan, A. (2001). Culture and systems of thought: Holistic versus analytic cognition. *Psychological Review, 108*, 291–310.

Nisbett, R. E., & Ross, L. (1980). *Human inference: Strategies and shortcomings of social judgment*. Englewood Cliffs, NJ: Prentice-Hall.

Olson, R. K. (1999). Genes, environment, and reading disabilities. In R. J. Sternberg & L. Spear-Swerling (Eds.), *Perspectives on learning disabilities* (pp. 3–21). New York: Westview/HarperCollins.

Olson, R. K., & Forsberg, H. (1993). Disabled and normal readers' eye movements in reading and nonreading tasks. In D. M. Willows, R. Kruk, & E. Corcos (Eds.), *Visual processes in reading and reading disabilities* (pp. 377–391). Hillsdale, NJ: Erlbaum.

Orkin, M. (2000). *What are the odds? Chance in eveyday life*. New York: W. H. Freeman.

Page, E., & Keith, T. (1981). Effects of U.S. private schools: A technical analysis of two recent claims. *Educational Researcher, 10*, 7–17.

Paik, H., & Comstock, G. (1994). The effects of television violence on antisocial behavior: A meta-analysis. *Communications Research, 21*, 516–546.

Paloutzian, R. F. (1983). *Invitation to the psychology of religion*. Glenview, IL: Scott, Foresman.

Paul, A. M. (2001, March/April 2001). Self-help: Shattering the myths. *Psychology Today*, pp. 62–68.

Paul, G. L. (1966). *Insight vs. desensitization in psychotherapy*. Stanford, CA: Stanford University Press.

Paul, G. L. (1967). Insight vs. desensitization in psychotherapy two years after termination. *Journal of Consulting Psychology, 31*, 333–348.

Paulos, J. A. (1988). *Innumeracy: Mathematical illiteracy and its consequences*. New York: Vintage Books.

Pearl, D., Bouthilet, L., & Lazar, J. (Eds.). (1982). *Television and behavior: Ten years of scientific progress and implications for the eighties*. Rockville, MD: National Institute of Mental Health.

Peng, K., & Nisbett, R. E. (1999). Culture, dialectics, and reasoning about contradiction. *American Psychologist, 54*, 741–754.

Pennington, B. F., & Ozonoff, S. (1996). Executive functions and developmental psychopathology. *Journal of Child Psychology and Psychiatry, 37*, 51–87.

Pennock, R. (1999). *Tower of Babel: The new creationism*. Cambridge, MA: MIT Press.

Perez, J. E. (1999). Clients deserve empirically supported treatments, not romanticism. *American Psychologist, 54*, 205–206.

Peterson, D. R. (1995). The reflective educator. *American Psychologist, 50*, 975–983.

Pettit, G. S., Bates, J. E., Dodge, K. A., & Meece, D. W. (1999). The impact of after-school peer contact on early adolescent externalizing problems is moderated by parental monitoring, perceived neighborhood safety, and prior adjustment. *Child Development, 70*, 768–778.

Pezdek, K., & Banks, W. P. (Eds.). (1996). *The recovered memory/false memory debate*. San Diego: Academic Press.

Pezdek, K., & Hodge, D. (1999). Planting false childhood memories in children: The role of event plausibility. *Child Development, 70*, 887–895.

Pfeifer, R., & Scheier, C. (1999). *Understanding intelligence*. Cambridge, MA: MIT Press.

Piattelli-Palmarini, M. (1994). *Inevitable illusions: How mistakes of reason rule our minds*. New York: John Wiley.

Pigliucci, M. (2002). Hypothesis testing and the nature of skeptical investigations. *Skeptical Inquirer, 26*(6), 27–48.

Pinker, S. (1997). *How the mind works*. New York: W. W. Norton.

Pinker, S. (2002). *The blank slate: The modern denial of human nature*. New York: Viking.

Piper, A. (1998). Multiple personality disorder: Witchcraft survives in the twentieth century. *Skeptical Inquirer, 22*(3), 44–50.

Poon, L. W., Rubin, D. C., & Wilson, B. A. (Eds.). (1989). *Everyday cognition in adulthood and late life*. Cambridge, England: Cambridge University Press.

Popper, K. R. (1959). *The logic of scientific discovery*. New York: Harper & Row.

Popper, K. R. (1963). *Conjectures and refutations*. New York: Harper & Row.
Popper, K. R. (1972). *Objective knowledge*. Oxford, England: Oxford University Press.
Popper, K. R. (1976). *Unended quest: An intellectual biography*. La Salle, IL: Open Court.
Postman, N. (1988). *Conscientious Objections*. New York: Vintage Books.
Poundstone, W. (1999). *Carl Sagan: A life in the cosmos*. New York: Henry Holt and Co.
Powell, B. (1993, December). Sloppy reasoning, misused data. *Phi Delta Kappan, 75*(4), 283, 352.
Powell, B., & Steelman, L. C. (1996). Bewitched, bothered, and bewildering: The use and misuse of state SAT and ACT scores. *Harvard Educational Review, 66*, 27–59.
Pressley, M. (2002). *Reading instruction that works: The case for balanced teaching* (2nd ed.). New York: Guilford Press.
Prioleau, L., Murdock, M., & Brody, N. (1983). An analysis of psychotherapy versus placebo studies. *Behavioral and Brain Sciences, 6*, 275–310.
Prior, M. P., & Cummins, R. A. (1992). Questions about facilitated communication and autism. *Journal of Autism and Developmental Disorders, 22*, 331–338.
Proctor, R. W., & Capaldi, E. J. (2001). Empirical evaluation and justification of methodologies. *Psychological Bulletin, 127*, 759–772.
Rabbitt, P. (1993). Does it all go together when it goes? *Quarterly Journal of Experimental Psychology, 46A*, 385–434.
Radford, B. (1999). The ten-percent myth. *Skeptical Inquirer, 23*(2), 52–53.
Ramey, S. L. (1999). Head Start and preschool education. *American Psychologist, 54*, 344–346.
Randi, J. (1983). The project alpha experiment: 1. The first two years. *Skeptical Inquirer, 7*, 24–33.
Randi, J. (1986). *Flim-flam*. Buffalo, NY: Prometheus Books.
Randi, J. (1987). *The faith healers*. Buffalo, NY: Prometheus Books.
Randi, J. (1995). *An encyclopedia of claims, frauds, and hoaxes of the occult and supernatural exposed by James Randi*. New York: St. Martin's Press.
Raymo, C. (1999). *Skeptics and true believers*. Toronto: Doubleday Canada.
Rayner, K. (1998). Eye movements in reading and information processing: 20 Years of research. *Psychological Bulletin, 124*, 372–422.
Rayner, K., Foorman, B. R., Perfetti, C. A., Pesetsky, D., & Seidenberg, M. S. (2001). How psychological science informs the teaching of reading. *Psychological Science in the Public Interest, 2*, 31–74.
Rayner, K., Foorman, B. R., Perfetti, C. A., Pesetsky, D., & Seidenberg, M. S. (2002, March). How should reading be taught? *Scientific American, 286*(3), 84–91.
Redding, R. E. (1998). *How common-sense psychology can inform law and psycholegal research*. University of Chicago Law School Roundtable, 5, 107–142.
Redelmeier, D. A., & Tibshirani, R. J. (1997). Association between cellular-telephone calls and motor vehicle collisions. *The New England Journal of Medicine, 336*(7), 453–458.
Redelmeier, D. A., & Tibshirani, R. J. (2001). Car phones and car crashes: Some popular misconceptions. *Canadian Medical Association Journal, 164*(11), 1581–1582.
Ridley, M. (1996). *Evolution* (2nd ed.). Cambridge, MA: Blackwell Science.
Ridley, M. (1999). *Genome: The autobiography of a species in 23 chapters*. New York: HarperCollins.
Rind, B., Tromovitch, P., & Bauserman, R. (2001). The validity and appropriateness of methods, analyses, and conclusions in Rind et al. (1998): A rebuttal of victimological critique from Ondersma et al. (2001) and Dallam et al. (2001). *Psychological Bulletin, 127*, 734–758.
Roberts, A. H., Kewman, D. G., Mercier, L., & Hovell, M. (1993). The power of nonspecific effects in healing: Implications for psychological and biological treatments. *Clinical Psychology Review, 13*, 375–391.
Robins, R. W., & Craik, K. H. (1994). A more appropriate test of the Kuhnian displacement thesis. *American Psychologist, 49*, 815–816.
Robins, R. W., Gosling, S. D., & Craik, K. H. (1999). An empirical analysis of trends in psychology. *American Psychologist, 54*, 117–128.

Robins, R. W.,Gosling, S. D., & Craik, K. H. (2000). Trends in psychology: An empirical issue. *American Psychologist, 55,* 276–277.

Robinson, D. N. (2001). Sigmund Koch—Philosophically speaking. *American Psychologist, 56,* 420–424.

Robinson, W. S. (1992). *Computers, minds, and robots.* Philadelphia: Temple University Press.

Rosenthal, R. (1990). How are we doing in soft psychology? *American Psychologist, 46,* 775–776.

Ross, L., & Nisbett, R. E. (1991). *The person and the situation: Perspectives of social psychology.* Philadelphia: Temple University Press.

Rotton, J., & Kelly, I. W. (1985). Much ado about the full moon. *Psychological Bulletin, 97,* 286–306.

Rowe, D. C., Vazsonyi, A. T., & Flannery, D. J. (1994). No more than skin deep: Ethnic and racial similarity in developmental process. *Psychological Review, 101,* 396–413.

Rowe, W. (1993). Psychic detectives: A critical examination. *Skeptical Inquirer, 17,* 159–165.

Rozin, P., Lowery, L., Imada, S., & Haidt, J. (1999). The CAD Triad Hypothesis: A mapping between three moral emotions (contempt, anger, disgust) and three moral codes (community, autonomy, divinity). *Journal of Personality and Social Psychology, 76,* 574–586.

Ruggiero, V. (2000). Bad attitude: Confronting the views that hinder students' learning. *American Educator, 24*(2), 10–15.

Ruscio, J. (2002, March). The emptiness of holism. *Skeptical Inquirer, 26*(2), 46–50.

Ruscio, J. (2000). Risky business: Vividness, availability, and the media paradox. *Skeptical Inquirer, 24*(2), 22–26.

Ruse, M. (1999). *Mystery of mysteries: Is evolution a social construction?* Cambridge, MA: Harvard University Press.

Russo, E. (2002, December 9). The biological basis of the placebo effect. *The Scientist, 16*(24), 30–31.

Russo, F. (1999, May). The clinical-trials bottleneck. *The Atlantic Monthly,* pp. 30–36.

Rutter, M. (1979). Maternal deprivation, 1972–1978: New findings, new concepts, and new approaches. *Child Development, 50,* 283–305.

Rzewnicki, R., & Forgays, D. G. (1987). Recidivism and self-cure of smoking and obesity: An attempt to replicate. *American Psychologist, 42,* 97–100.

Sagan, C. (1996). *The demon-haunted world: Science as a candle in the dark.* New York: Random House.

Salthouse, T. A. (1996). Constraints on theories of cognitive aging. *Psychonomic Bulletin & Review, 3,* 287–299.

Salthouse, T. A., & Miles, J. D. (2002). Aging and time-sharing aspects of executive control. *Memory & Cognition, 30,* 572–582.

Sandoval, J., & Miille, M. (1980). Accuracy of judgments of WISC-R item difficulty for minority groups. *Journal of Consulting and Clinical Psychology, 48,* 249–253.

Santrock, J. W., Minnett, A. M., & Campbell, B. D. (1994). *The authoritative guide to self-help books.* New York: Guilford Press.

Sawyer, J. (1996). Measurement and prediction, clinical and statistical. *Psychological Bulletin, 66,* 178–200.

Scarborough, D., & Sternberg, S. (Eds.). (1998). *Methods, models, and conceptual issues: An invitation to cognitive science* (Vol. 4). Cambridge, MA: MIT Press.

Schachter, S. (1982). Recidivism and self-cure of smoking and obesity. *American Psychologist, 37,* 436–444.

Schraw, G., & Nietfeld, J. (1998). A further test of the general monitoring skill hypothesis. *Journal of Educational Psychology, 90,* 236–248.

Scott, S. (1999, January 2). Risking all on alternative cancer therapies. *National Post (Toronto),* p. B1.

Seligmann, J., & Chideya, F. (1992, September 21). Horror story or big hoax? *Newsweek,* p. 75.

Senechal, M., & LeFevre, J. (2002). Parental involvement in the development of children's reading skill: A five-year longitudinal study. *Child Development, 73,* 445–460.

Shaffer, L. (1981). The growth and limits of recipe knowledge. *Journal of Mind and Behavior, 2,* 71–83.

Shane, H. (1993, June). FC: Facilitated or "factitious" communication. *Communicating Together, 11*(2), 11–13.

Shanks, D. R. (1995). Is human learning rational? *Quarterly Journal of Experimental Psychology, 48A,* 257–279.
Shapiro, A. (1960). A contribution to a history of the placebo effect. *Behavioral Science, 5,* 109–135.
Shapiro, A., Shapiro, E., Bruun, R., & Sweet, R. (1978). *Gilles de la Tourette syndrome.* New York: Raven Press.
Share, D. L., & Stanovich, K. E. (1995). Cognitive processes in early reading development: Accommodating individual differences into a model of acquisition. *Issues in Education: Contributions from Educational Psychology, 1,* 1–57.
Shaywitz, S. E. (1996). Dyslexia. *Scientific American, 275*(5), 98–104.
Shepard, R. (1983). "Idealized" figures in textbooks versus psychology as an empirical science. *American Psychologist, 38,* 855.
Shermer, M. (1997). *Why people believe weird things.* New York: W.H. Freeman.
Shontz, F., & Green, P. (1992). Trends in research on the Rorschach: Review and recommendations. *Applied and Preventive Psychology, 1,* 149–156.
Siegert, R. J., & Ward, T. (2002). Clinical psychology and evolutionary psychology: Toward a dialogue. *Review of General Psychology, 6,* 235–259.
Sigman, M., & Ruskin, E. (1999). Continuity and change in the social competence of children with autism, Down syndrome, and developmental delays. *Monographs of the Society for Research in Child Development, 64*(Serial No. 256), 1–114.
Simmons, R., Burgeson, R., Carlton-Ford, S., & Blyth, D. (1987). The impact of cumulative change in early adolescence. *Child Development, 58,* 1220–1234.
Simpson, R. L., & Myles, B. S. (1995). Effectivenesss of facilitated communication with children and youth with autism. *Journal of Special Education, 28,* 424–439.
Sivak, M., & Flannagan, M. J. (2003). Flying and driving after the September 11 attacks. *American Scientist, 91,* 6–7.
Slemrod, J., & Bakija, J. (1996). *Taxing ourselves: A citizen's guide to the great debate over tax reform.* Cambridge, MA: The MIT Press.
Smith, M., & Belcher, R. (1993). Facilitated communication with adults with autism. *Journal of Autism and Developmental Disorders, 23,* 175–183.
Smith, M. L., Glass, G. V., & Miller, T. I. (1980). *The benefits of psychotherapy.* Baltimore: Johns Hopkins University Press.
Smith, R. E., Wheeler, G., & Diener, E. (1975). Faith without works. *Journal of Applied Social Psychology, 5,* 320–330.
Snow, C. E., Burns, M. S., & Griffin, P. (Eds.). (1998). *Preventing reading difficulties in young children.* Washington, DC: National Academy Press.
Solso, R., & Massaro, D. W. (1996). *The science of the mind.* New York: Oxford University Press.
Spanos, N. P. (1996). *Multiple identities and false memories: A sociocognitive perspective.* Washington, DC: American Psychological Association.
Spitz, H. H. (1997). *Nonconscious movements: From mystical messages to facilitated communication.* Mahwah, NJ: Erlbaum.
Stankov, L., & Spilsbury, G. (1978). The measurement of auditory abilities of blind, partially sighted, and sighted children. *Applied Psychological Measurement, 2,* 491–503.
Stanovich, K. E. (1989). Implicit philosophies of mind: The dualism scale and its relationships with religiosity and belief in extrasensory perception. *Journal of Psychology, 123,* 5–23.
Stanovich, K. E. (1999). *Who is rational? Studies of individual differences in reasoning.* Mahwah, NJ: Erlbaum.
Stanovich, K. E. (2000). *Progress in understanding reading.* New York: Guilford Press.
Stanovich, K. E. (in press). *Rationality and evolution: The robot's rebellion in the age of Darwin.* Chicago: University of Chicago Press.
Stanovich, K. E., & West, R. F. (1998). Who uses base rates and P(D/~H)? An analysis of individual differences. *Memory & Cognition, 28,* 161–179.
Steen, L. A. (1990). Numeracy. *Daedalus, 119,* 211–231.
Steinberg, L., Brown, B. B., & Dornbusch, S. M. (1996). *Beyond the classroom: Why school reform has failed and what parents need to do.* New York: Simon & Schuster.
Steinberg, L., Fegley, S., & Dornbusch, S. M. (1993). Negative impact of part-time

work on adolescent adjustment: Evidence from a longitudinal study. *Developmental Psychology, 29,* 171–180.

Sternberg, R. J. (Ed.). (1999). *The nature of cognition.* Cambridge, MA: MIT Press.

Sternberg, R. J. (Ed.). (2000). *Handbook of intelligence.* New York: Cambridge University Press.

Sternberg, R. J. (Ed.). (2002). *Why smart people can be so stupid.* New Haven, CT: Yale University Press.

Sternberg, R. J., & Kaufman, J. C. (1998). Human abilities. *Annual Review of Psychology, 49,* 479–502.

Sternberg, R. J., & Spear-Swerling, L. (Eds.). (1999). *Perspectives on learning disabilities* (pp. 3–21). New York: Westview/HarperCollins.

Sternberg, R. J., & Wagner, R. K. (Eds.). (1994). *Mind in context.* Cambridge, England: Cambridge University Press.

Stevenson, H., Stigler, J., Lee, S., Lucker, G., Kitamura, S., & Hsu, C. (1985). Cognitive performance and academic achievement of Japanese, Chinese, and American children. *Child Development, 56,* 718–734.

Stout, M. (2000). *The feel-good curriculum: The dumbing down of America's kids in the name of self-esteem.* Cambridge, MA: Perseus Publishing.

Strupp, H. H. (1989). Psychotherapy: Can the practitioner learn from the researcher? *American Psychologist, 44,* 717–724.

Strayer, D. L., & Johnston, W. A. (2001). Driven to distraction: Dual-task studies of simulated driving and and conversing on a cellular telephone. *Pschological Science, 12,* 462–466.

Sutherland, S. (1992). *Irrationality: The enemy within.* London: Constable.

Svenson, O. (1981). Are we all less risky and more skillful than our fellow drivers? *Acta Psychologica, 47,* 143–148.

Swets, J. A., Dawes, R. M., & Monahan, J. (2000). Psychological science can improve diagnostic decisions. *Psychological Science in the Public Interest, 1,* 1–26.

Taleb, N. N. (2001). *Fooled by randomness.* New York: Texere.

Tassoni, C. J. (1996). Representativeness in the market for bets on National Football League games. *Journal of Behavioral Decision Making, 9,* 115–124.

Taube, K. T., & Linden, K. W. (1989). State mean SAT score as a function of participation rate and other educational and demographic variables. *Applied Measurement in Education, 2,* 143–159.

Teigen, K. H. (1986). Old truths or fresh insights? A study of students' evaluations of proverbs. *British Journal of Social Psychology, 25,* 43–49.

Thaler, R. H. (1992). *The winner's curse: Paradoxes and anomalies of economic life.* New York: Free Press.

Thomas, G., Alexander, K., & Eckland, B. (1979). Access to higher education: The importance of race, sex, social class, and academic credentials. *School Review, 87,* 133–156.

Thornton, E. (1986). *The Freudian fallacy.* London: Paladin Books.

Tooby, J. (2002, October 6). The greatest Englishman since Newton. *New York Times Book Review,* p. 12.

Turing, A. (1950). Computing machinery and intelligence. *Mind, 59,* 433–460.

Turner, C., Simons, L., Berkowitz, L., & Frodi, A. (1977). The stimulating and inhibiting effects of weapons on aggressive behavior. *Aggressive Behavior, 3,* 355–378.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science, 185,* 1124–1131.

Twachtman-Cullen, D. (1997). *A passion to believe.* Boulder, CO: Westview.

U.S. Congress House Select Committee on Aging. (1984, May 31). *Quackery: A $10 billion scandal.* Washington, DC: U.S. Government Printing Office.

University of California, Berkeley. (1991, January). The 18-year gap. *Berkeley Wellness Letter,* p. 2.

Updegrave, W. L. (1995, August). Why funds don't do better. *Money Magazine,* pp. 58–65.

Vallone, R., Griffin, D. W., Lin, S., & Ross, L. (1990). Overconfident prediction of future actions and outcomes by self and others. *Journal of Personality and Social Psychology, 58,* 582–592.

Vaughan, E. D. (1977). Misconceptions about psychology among introductory psychology students. *Teaching of Psychology, 4,* 138–141.

Violato, C. (1984). Effects of Canadianization of American-biased items on the WAIS

and WAIS-R information subtests. *Canadian Journal of Behavioral Science, 16,* 36–41.

Violato, C. (1986). Canadian versions of the information subtests of the Wechsler tests of intelligence. *Canadian Psychology, 27,* 69–74.

Wachs, T. D. (2000). *Necessary but not sufficient: The respective roles of single and multiple influences on individual development.* Washington, DC: American Psychological Association.

Wagenaar, W. A. (1988). *Paradoxes of gambling behavior.* Hove, England: Erlbaum.

Wagenaar, W. A., & Keren, G. (1986). The seat belt paradox: Effect of adopted roles on information seeking. *Organizational Behavior and Human Decision Processes, 38,* 1–6.

Wainer, H. (1989). Eelworms, bullet holes, and Geraldine Ferraro: Some problems with statistical adjustment and some solutions. *Journal of Educational Statistics, 14,* 121–140.

Wainer, H. (1993). Does spending money on education help? A reaction to the Heritage Foundation and the Wall Street Journal. *Educational Researcher, 22*(9), 22–24.

Wainer, H. (1999). The most dangerous profession: A note on nonsampling error. *Psychological Methods, 4,* 250–256.

Wampold, B. E., Mondin, G. W., Moody, M., Stich, F., Benson, K., & Ahn, H. (1997). A meta-analysis of outcome studies comparing bona fide psychotherapies: Empirically, "all must have prizes." *Psychological Bulletin, 122,* 203–215.

Watters, E., & Ofshe, R. (1999). *Therapy's delusions: The myth of the unconscious and the exploitation of the walking worried.* New York: Scribners.

Welch, D. A. (2002). *Decisions, decisions: The art of effective decision making.* Amherst, NY: Prometheus Books.

Wheeler, D. L., Jacobson, J. W., Paglieri, R. A., & Schwartz, A. A. (1993). An experimental assessment of facilitated communication. *Mental Retardation, 31,* 49–60.

Wickens, C. D. (1992). *Engineering psychology and human performance* (2nd ed.). New York: HarperCollins.

Winter, M. G., North, C., & Sugar, L. A. (2001). Psychologists' response to criticisms about research based on undergraduate participants: A developmental perspective. *Canadian Psychology, 42,* 216–225.

Wilson, E. O. (1998). *Consilience: The unity of knowledge.* New York: Knopf.

Wilson, R. A., & Keil, F. C. (Eds.). (1999). *The MIT Encyclopedia of the cognitive sciences.* Cambridge, MA: MIT Press.

Wilson, T. D., & Brekke, N. (1994). Mental contamination and mental correction: Unwanted influences on judgments and evaluations. *Psychological Bulletin, 116,* 117–142.

Wiseman, R., Beloff, J., & Morris, R. L. (1996). Testing the ESP claims of SORRAT. *Skeptical Inquirer, 20*(5), 45–61.

Wolfle, L. M. (1987). Enduring cognitive effects of public and private schools. *Educational Researcher, 16,* 5–11.

Wood, J. M., Nezworski, T., & Stejskal, W. J. (1996). The comprehensive system for the Rorschach: A critical examination. *Psychological Science, 7,* 3–10.

Wood, W., Wong, F. Y., & Chachere, J. G. (1991). Effects of media violence on viewers' aggression in unconstrained social interaction. *Psychological Bulletin, 109,* 371–383.

Woodcock, R. W. (1987). *Woodcock Reading Mastery Tests—Revised.* Circle Pines, MN: American Guidance Service.

Woolley, B. (2000). *The bride of science: Romance, reason, and Byron's daughter.* New York: McGraw-Hill.

Wright, L. (1988). The Type A behavior pattern and coronary artery disease. *American Psychologist, 43,* 2–14.

Wright, R. (1988). *Three scientists and their gods.* New York: Harper & Row.

Wright, R. (1999, December 13). The accidental creationist. *The New Yorker,* pp. 56–65.

Wrightsman, L. S. (1999). *Judicial decision making: Is psychology relevant?* New York: Kluwer Academic/Plenum.

Yates, J. F., Lee, J., & Bush, J. G. (1997). General knowledge overconfidence: Cross-national variations, response style, and "reality." *Organizational Behavior and Human Decision Processes, 70,* 87–94.

Yates, J. F., Lee, J., & Shinotsuka, H. (1996). Beliefs about overconfidence, including its cross-national variation. *Organizational Behavior and Human Decision Processes, 65,* 138–147.

Yirmiya, N., Erel, O., Shaked, M., & Solomonica-Levi, D. (1998). Meta-analysis comparing theory of mind abilities of individuals with autism, individuals with mental retardation, and normally developing individuals. *Psychological Bulletin, 124,* 283–307.

Zechmeister, J. S., & Zechmeister, E. B. (2000). Introductory textbooks and psychology's core concepts. *Teaching of Psychology, 27,* 6–11.

Zill, N., & Winglee, M. (1990). *Who reads literature?* Cabin John, MD: Seven Locks Press.

Zweig, J. (1998, July). Here's how to use the news and tune out the noise. *Money Magazine,* pp. 63–64.

Zweig, J. (2001, May). Do you sabotage yourself? *Money Magazine,* pp. 75–78.

Zweig, J. (2002, October). Are you wired for wealth? *Money Magazine,* pp. 75–83.

國家圖書館出版品預行編目資料

這才是心理學！：教你如何以科學方法，培養批判性思考的能力／ Keith E. Stanovich 作；楊中芳等譯. -- 初版. -- 臺北市：遠流, 2005[民 94]

面；　公分. --（大眾心理學叢書；300)

參考書目：面

譯自： How to think straight about psychology, 7th ed.

ISBN 957-32-5490-5（平裝）

1. 心理學 - 研究方法　2. 大眾傳播 - 心理方面

170.12　　94003886

在大衆媒體上，在書店的自助類書籍中，你一定經常看到一些「見人所未見」的所謂心理學家大放厥詞。小心上當！這些人大多是僞心理學家——因爲他們不科學，因爲他們宣揚的不是眞正的心理學。

那，什麼是心理學？本書作者指出，心理學是一套可以用來解釋、預測和控制人的行爲與意識的知識體系。眞正科學的心理學是以實徵研究爲本，用可操作的程序定義理論中的概念，透過公開蒐集的資料、系統實徵的方法檢驗理論，允許其他科學家重複這些實驗並提出批評，以控制和操作的邏輯從事實驗研究，以統計爲工具進行相關與因果研究的嚴謹評鑑。正是透過這樣的去僞存眞，我們才得以看清心理學的眞正面貌。

本書是爲有心應用心理學資訊的人所寫，適用於初學心理學的學生，也適用於透過大衆媒體對心理學知識有些了解、又想知道如何評價這些資訊之合理性的廣大讀者。從本書歷經七次修訂改版可知，有這種需求的讀者爲數可觀。

以科學的精神、用科學研究的方法與工具探討人類的行爲，豐富我們對自身及環境的理解——這才是心理學！

只要讀者捧起書，一定會為它精采的內容所吸引，讀完之後也一定會對「人」的複雜有所感觸，以後更會對心理學的各項報導，產生自發性的批判性思考。不論讀者是心理學或其他學科的專業，都將因為這本書而開始對認真研究真正的心理學的人有所尊敬。

——中央研究院副院長 曾志朗

ISBN 957-32-5490-5
〔170〕
00320
9 789573 254904
A3300 NT$320